

UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
MESTRADO EM ADMINISTRAÇÃO

Péricles Fontanella

**Associações de Compra em Supermercado utilizando o
Data Mining**

CURITIBA
2010

UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
MESTRADO EM ADMINISTRAÇÃO

**Associações de Compra em Supermercado utilizando o
Data Mining**

Dissertação apresentada ao Centro de Pesquisa e Pós Graduação em Administração do Departamento de Administração da Universidade Federal da Paraná (CEPPAD/UFPR), para obtenção do título de Mestre em Administração.

Aluno: Péricles Fontanella

Orientador: Prof. Dr. Valter Afonso Vieira

CURITIBA
2010

TERMO DE APROVAÇÃO

Péricles Fontanella

“Associação de compra em supermercados utilizando o data mining”

APROVADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE NO PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO DA UNIVERSIDADE FEDERAL DO PARANÁ, PELA SEGUINTE BANCA EXAMINADORA:


Prof. Dr. Valter Afonso Vieira
(Orientador/UFPR)


Prof. Dr. José Eduardo Pécora Junior
(Examinador/UFPR)


Prof.ª Dr.ª Ana Maria Machado Toaldo
(Examinadora/UFPR)


Prof. Dr. Paulo de Paula Baptista
(Examinador/Puc/PR)

16 de dezembro de 2010

AGRADECIMENTOS

Ao Prof. Dr. Valter Afonso Vieira, amigo, incentivador e orientador, que nunca mediu esforços para a realização deste trabalho.

Aos amigos que tive oportunidade de fazer neste curso, em especial Alexandre (Urso), Ligia, Valmir Reich, Leandro.

À equipe do Supermercado Franzoni, que sempre estavam prontos para auxiliar com informações relevantes para a realização deste trabalho.

Ao Sr. João Marcio Junior, que através de seu software viabilizou a coleta e análise dos dados.

À minha mãe Irma e ao meu pai Geraldo Fontanella.

Aos examinadores da Banca: Dra. Ana Maria Toaldo, Dr. Paulo Baptista e Dr. Jose Eduardo Pécora Junior.

Em especial à minha esposa Cyntia e ao meu filho Murilo, que sempre tiveram paciência e compreensão nos momentos que estive ausente.

E à Deus por sempre guiar os meus caminhos.

Muito Obrigado à todos!

LISTA DE FIGURAS

Figura 1: Market Share de ERPs no Brasil em 2008 Fonte FGV.....	14
Figura 2: Estrutura típica de funcionamento de um sistema ERP.	14
Figura 3: Modelos lógicos de banco de dados.	17
Figura 4: Arquitetura Data Warehouse.....	18
Figura 5: Processamento <i>Data Mining</i>	23
Figura 6: Etapas do processo de Mineração de Dados.	29
Figura 7: Busca por informações em sistemas convencionais (SIG, OLAP, Query, etc.)	34
Figura 8: Busca por informações em sistemas de mineração de dados	34
Figura 9: Modelo Integrado de Gerenciamento pelo Valor do Cliente no Tempo.....	44
Figura 10: Metodologia CRISP-DM	46
Figura 11: Passos para <i>Data Mining</i>	57
Figura 12: Etapas do processo de KDD utilizadas no estudo	64

LISTA DE FOTOS

Foto 1: Supermercado Franzoni & Cia Ltda.....	77
Foto 2: Supermercado Franzoni & Cia Ltda - setor hortifrutigranjeiros.....	77
Foto 3: Supermercado Franzoni & Cia Ltda - <i>check outs</i>	78
Foto 4: Supermercado Franzoni & Cia Ltda - estacionamento.....	78
Foto 5: Supermercado Franzoni & Cia Ltda - lanchonete.....	79
Foto 6: Mapa da cidade de Quedas do Iguaçu – indicando a localização do Supermercado.....	79
Foto 7: Combo I em exposição no supermercado.....	104
Foto 8: Combo II em exposição no supermercado.....	106
Foto 9: Combo III em exposição no supermercado.....	108
Foto 10: Combo IV em exposição no supermercado.....	110

LISTA DE QUADROS

Quadro 1: Papel desenvolvido pelos varejistas e atacadistas no mercado.....	7
Quadro 2: Tipos de atacadistas.....	7
Quadro 3: Fases e atividades genéricas do modelo CRISP-DM.....	47
Quadro 4: Fontes de Informações Disponibilizadas pela Empresa em Estudo	50
Quadro 5: Relação de Entrevistas Exploratórias com Empresas Usuárias de VoIP.....	52
Quadro 6: Diferença entre as informações adquiridas sobre os clientes.....	54
Quadro 7: Exemplo de base de dados, com compras realizadas.....	72
Quadro 8: Exemplo da base de dados extraídas do supermercado.....	74

LISTA DE TABELAS

Tabela 1: Atributos de serviços para supermercado	61
Tabela 2: Associações no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	81
Tabela 3: Associações no período bimestral de Janeiro-Fevereiro de 2009 com suporte relevantes.....	84
Tabela 4: Associações no período bimestral de Março-Abril de 2009 com suporte relevantes.....	85
Tabela 5: Associações no período bimestral de Maio-Junho de 2009 com suporte relevantes.....	86
Tabela 6: Associações no período bimestral de Julho-Agosto de 2009 com suporte relevantes.....	87
Tabela 7: Associações no período bimestral de Setembro-Outubro de 2009 com suporte relevantes.....	88
Tabela 8: Associações no período bimestral de Novembro-Dezembro de 2009 com suporte relevantes.....	88
Tabela 9: Produto Carne - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	90
Tabela 10: Produto Material de Limpeza - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	92
Tabela 11: Produto Refrigerante - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	93
Tabela 12: Produto Açúcar - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	94
Tabela 13: Produto Óleo - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	95
Tabela 14: Produto Sal - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	96
Tabela 15: Produto Farinha de Trigo - no período de Janeiro-Dezembro de 2009 com suporte relevantes.....	97
Tabela 16: Produto Tomate - no período de Janeiro-Dezembro de 2009 com	

suporte relevantes.....	98
Tabela 17: Produto Diversos - no período de Janeiro-Dezembro de 2009 com três itens associados e com suporte relevantes.....	100
Tabela 18: Produto Diversos - no período de Janeiro-Dezembro de 2009 com três itens associados e com confiança relevantes.....	101
Tabela 19: Análise Combo I - período de Maio de 2009 e Maio de 2010.....	103
Tabela 20: Análise Combo I - período de Maio-Junho de 2009 e Maio-Junho de 2010.....	103
Tabela 21: Análise Combo II - período de Maio de 2009 e Maio de 2010.....	105
Tabela 22: Análise Combo II - período de Maio-Junho de 2009 e Maio-Junho de 2010.....	105
Tabela 23: Análise Combo III - período de Maio de 2009 e Maio de 2010.....	106
Tabela 24: Análise Combo III - período de Maio-Junho de 2009 e Maio-Junho de 2010.....	107
Tabela 25: Análise Combo IV - período de Maio de 2009 e Maio de 2010.....	108
Tabela 26: Análise Combo IV - período de Maio de 2009 e Maio de 2010.....	109
Tabela 27: Resumo de todas operações realizadas e variações - período de Maio de 2009 e Maio de 2010.....	111
Tabela 28: Resumo de todas operações realizadas e variações - período de Maio-Junho de 2009 e Maio-Junho de 2010.....	114

LISTA DE SIGLAS

OLAP	<i>On Line Analytical Processing</i>
QUERY	<i>Computer languages used to make queries into databases and information systems</i>
CRM	<i>Consumer Relationship Management</i>
ERP	<i>Enterprise Resources Planning</i>
CLV	<i>Customer Lifetime Value</i>
SGBD	Sistema de gerenciamento de banco de dados.
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
ETL	<i>Extract, Transform and load</i>
OLTP	<i>On-line transaction processing</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IPARDES	Instituto Paranaense para o Desenvolvimento Econômico e Social
MTE	Ministério do Trabalho e Emprego
CAD/CAM	<i>Computer Aided Design/Computer Aided Manufacturing</i>
MBR	<i>Memory Based reasoning</i>
CART	<i>Classification and Regression Trees</i>
CHAID	<i>Chi-squared Automatic Interaction Detection</i>
LDL++	<i>Database System</i>
POS	Ponto de venda

SUMÁRIO

1 INTRODUÇÃO	1
1.1 Formulação do Problema	3
1.2 Objetivos	3
1.2.1 Objetivo Geral	3
1.2.2 Objetivos Específicos	3
1.3 Relevância Teórica.....	4
1.4 Relevância Prática	4
2 FUNDAMENTAÇÃO TEÓRICA	6
2.1 Varejo	6
2.1.1 Supermercado.....	9
2.2 <i>Enterprise Resources Planning</i> - ERP	10
2.3 Banco de Dados.....	16
2.4 <i>Data Mining</i>	21
2.5 Estudos Anteriores sobre <i>Data Mining</i>	63
3 METODOLOGIA	68
3.1 Pesquisa Descritiva	68
3.2. Bando de Dados.....	68
3.3 Período de Tempo Longitudinal	69
3.4 Amostra	69
3.5 Análise de Dados: Regras de Associação.....	70
3.6 Varejista	76
4 ANÁLISE DOS DOS DADOS.....	80
4.1 Total de Produtos no Supermercado – Análise Anual	80
4.2 Total de Produtos no Supermercado – Análise Bimestral	83
4.3. Segmentos de Produtos no Supermercado - Anual	89
4.4 Quatro Combos Sugeridos ao Varejista	101
5 CONSIDERAÇÕES FINAIS.....	117
5.1 Sugestões para Pesquisas Futuras.....	120
5.2 Sugestões para o Supermercado	121
5.3 Limitações do Trabalho	122
REFERÊNCIAS.....	124

RESUMO

Esta pesquisa teve como objetivo identificar as principais associações de compras de produtos realizadas pelos consumidores em um supermercado de pequeno porte, na cidade de Quedas do Iguaçu, interior do Paraná, a partir do uso da ferramenta *Data Mining*, utilizando-se das regras de associação. O trabalho foi desenvolvido no Supermercado Franzoni, através da análise do banco de dados do período de Janeiro até Dezembro de 2009. Primeiro, os resultados encontrados demonstram que a ferramenta *Data Mining* ajuda a identificar determinados setores de maior volume de vendas e produtos de maior relevância dentro do *mix* de vendas da organização, como é o caso do setor hortifrutigranjeiro (batata monaliza e banana catura). Segundo, após a análise foram sugeridos os combos de vendas e, após implantados, auxiliaram na verificação e validação das associações descobertas na fase anterior. Assim sendo, os combos confirmaram que a conversão dos dados garimpados em informações possibilitam alavancar vendas dos produtos Limpol Multi Uso 500ml, Fruta Citr Q-Boa, e Pêssego em Calda 400g. A presente pesquisa contribui academicamente demonstrando que a associação de compra auxilia o gestor em problemas práticos, trazendo novas oportunidades para serem exploradas via comunicação de marketing.

Palavras-Chave: Mineração de Dados. Data Mining. Regras de Associação. Supermercado.

ABSTRACT

This research had as objective to identify the main purchase associations of products in a small-sized supermarket, at Quedas do Iguaçu city, state of Paraná, from the use of the *Data Mining* tool, using of the association rules. The work was developed at Franzoni Supermarket, through the analysis of the data base in the period from January to December of 2009. The joined results demonstrate that tools as Data Mining helps to identify determined sectors and products of relevance of the mix of sales in the organization, as it is the case of the agricultural products sector (monalisa potato and catura banana); the suggested combos and implanted to assist in the verification and validation of the discovered associations had confirmed that the conversion of the data in information makes possible to leverage the sales (Limp Multi Uso 500ml Fruta Citr, Q-Boa 8400% e Pêssego em Calda 400g. *Metades Cantu 5400%*); also suggests, that seasonality can indicate better periods of commercialization and combos composition. The present research aims contribute academically bringing new challenges and opportunities to administration research realization by Data Mining application.

Keywords: *Data Mining*. Rules of Association. Supermarket.

1 INTRODUÇÃO

O mundo empresarial está passando por uma série de transformações nas últimas décadas, sendo a informação a protagonista neste cenário, uma vez que ela é ensejadora de alguns pontos estratégicos cruciais. Estes pontos estratégicos cruciais são: facilita os processos internos, possibilita inovar saindo à frente dos concorrentes, é uma das responsáveis pela melhoria da qualidade e diversifica produtos e serviços. De acordo com Lesca e Almeida (1994, p.66), “é fácil observar que a **informação** estratégica de qualidade, relevante, precisa, clara, consistente e oportuna possui um valor significativo para as organizações, podendo ser aplicada em diferentes contextos”.

Almeida (1994) destaca que uma gestão de qualidade permite às organizações a possibilidade de aproveitar e explorar as diversas oportunidades no ambiente dinâmico dos negócios globalizados. Almeida (1994) também garante, a melhoria contínua das diversas atividades empresariais por meio da capacidade de processamento de informações, a qual auxilia as organizações a obterem uma gestão ágil, de qualidade e com previsão, facilitando o planejamento estratégico do setor.

O processo de produzir informações gerenciais manualmente (via análise de relatórios), no entanto, é na maioria das vezes dispendioso, demorado e cansativo, haja visto a quantidade de dados a serem coletados e trabalhados para que seja viável uma análise. Assim, muitas empresas têm optado pela criação de um ambiente que traga fluidez a esse processo. Desta forma, pode-se dizer que a Tecnologia da Informação é uma das principais ferramentas de apoio para a tomada de decisão nas organizações.

Assim, com o advento tecnológico, e conseqüentemente com as mudanças frenéticas advindas da era do conhecimento, a informação passa a ser fator preponderante, devido à necessidade constante do aperfeiçoamento dos serviços e dos produtos da organização e até mesmo da sua estratégia empresarial. Para que esses processos sejam eficazes no auxílio da tomada de decisão é necessário que sejam partes constantes no planejamento estratégico.

A informação está presente em todos os lugares, e as empresas precisam de todas elas, porém, se faz necessário que estas informações sejam agrupadas, ordenadas e organizadas de forma a serem entendidas pelos clientes internos e

externos da empresa. As empresas necessitam de Sistema de Informação (SI) e Tecnologia da Informação (TI) que organizem essas informações para melhor serem utilizadas em suas atividades rotineiras e na tomada de decisão eficaz, ponto fundamental para um bom desenvolvimento e sucesso empresarial.

O SI, em integração com a TI, fornece uma ferramenta de análise de dados conhecida como *data mining*. O *data mining* é primordial para gerir toda essa quantidade de informação que as organizações possuem, permitindo melhor integridade e disponibilidade, facilitando a disseminação das informações, proporcionando melhor comunicação e poder de análise para os gestores.

Dentro do bojo do *data mining*, aparece como protagonista a técnica de mineração de dados, a qual tem como objetivo encontrar padrões de repetição em dados armazenados em um banco de dados.

Partindo por um prisma mais específico, há diversos tipos de tarefas de mineração, sendo as mais usuais as tarefas de associação, classificação e agrupamento. Na tarefa de associação, busca-se por exemplo, identificar itens que possam ser colocados juntos em um mesmo cesto, verificando assim, a possibilidade de uma pessoa adquirir determinados produtos associados. Através destas informações, a empresa pode aumentar sua capacidade de explorar as estratégias de vendas.

Considerando os aspectos citados, o presente estudo é orientado pela aplicação do *Data Mining* em supermercado, utilizando-se como métrica a regra de associação.

Como problema de pesquisa, tem-se: **Quais são as principais associações de compra encontradas em supermercado de pequeno porte, na cidade de Quedas do Iguaçu, no interior do Paraná, a partir do uso da ferramenta *Data Mining*?**

1.1 Formulação do Problema

Quais são as principais associações de compra encontradas em supermercado de pequeno porte, na cidade de Quedas do Iguaçu, no interior do Paraná, a partir do uso da ferramenta *Data Mining*?

1.2 Objetivos

1.2.1 Objetivo Geral

Identificar as principais associações de compra encontradas em supermercado de pequeno porte, em Quedas do Iguaçu, com o uso da ferramenta *Data Mining*.

1.2.2 Objetivos Específicos

- Identificar as associações de compra em supermercado de pequeno porte na cidade de Quedas do Iguaçu, bimestralmente e anualmente;
- Identificar as associações de compra por segmento de produtos. Tais como, por exemplo: carnes, bebidas, material de limpeza, etc;
- Propor combos de vendas de produtos (um vez que foram identificados), através da exposição em ponto de gôndolas, associando produtos correlacionados; e
- Avaliar o incremento nas vendas dos combos propostos por meio de uma análise antes e depois do volume de vendas.

1.3 Relevância Teórica

Com a evolução da informática, a sociedade do conhecimento foi se integrando a TI. A informação, assim, passa a ser preponderante para a sobrevivência organizacional dado o advento da globalização, tornando-se o computador o principal elemento das transformações sociais. Informações com qualidade e apresentadas em tempo hábil à tomada de decisão são de vital importância para as empresas modernas. O uso adequado dos recursos da TI garante a qualidade e a pontualidade das informações.

Os estudos relacionados à *Data Mining* mostram-se relevantes para a academia e para a prática da administração pois contribuem para a descoberta e aprimoramento de técnicas e métodos de mineração de dados, os quais transformam a volumosa quantidade de dados em informações úteis, facilitando e guiando gestores na tomada de decisão.

1.4 Relevância Prática

Algumas organizações ainda acreditam que dispor de TI se resume em adquirir computadores, *softwares* e outros equipamentos tecnológicos, disponibilizando-os para seus colaboradores. Assim, a empresa gasta tempo, recursos e dinheiro, mas a TI não se resume a este procedimento de aquisição computadores.

As informações estão presentes em todos os lugares, porém, é necessário que estas informações sejam agrupadas, ordenadas e organizadas de forma a serem entendidas pelos clientes internos e clientes externos da empresa. As empresas, portanto, necessitam de TIs que organizem as informações para melhor serem utilizadas em suas atividades rotineiras e na tomada de decisão eficaz, ponto fundamental para um bom desenvolvimento e sucesso empresarial.

Dentre as ferramentas tecnológicas que tem surgido no mercado, a Mineração de Dados tem auxiliado a garimpar informações valiosas e coerentes para a transformação e implementação de estratégias eficazes. De acordo com Barbosa (2007) a grande quantidade de dados e informações geradas diariamente nas organizações tem crescido muito, influenciando diretamente no processo de

tomada de decisões, pelo fato de que é preciso estar preparado para as possíveis mudanças e desafios do presente e também perante às oportunidades do futuro.

Ademais, o mercado atual está agressivo e, conforme pontua Gonçalves, (1999), utilizar ferramentas, técnicas e tecnologias para trabalhar com a informação é um fator de influência no aumento da competitividade organizacional. Dessa maneira, conhecer as ferramentas que possam ser utilizadas para este fim é de suma importância.

Nesta perspectiva, para aceitar e implementar qualquer tecnologia deve contar com um detalhado estudo não somente dos benefícios, mas dos riscos que as mesmas possam porventura apresentar. Demonstrar as dificuldades de comportamento no momento da adoção do *Data Mining* para minerar dados pode auxiliar as empresas a estabelecer estratégias de maneira mais eficiente.

A presente pesquisa utiliza a ferramenta de *Data Mining*, através da tarefa regras de associação, extraíndo informações do banco de dados que antes eram desconhecidos e não agregavam valor, gerando maiores possibilidades na formulação de estratégias de marketing, no gerenciamento de investimentos, e em melhor gestão do compostos de marketing. Tais estratégias uma vez implementadas geram, por consequência, melhores produtos e serviços ao consumidor, elevando assim o nível de competitividade organizacional.

Destarte, esta pesquisa é importante aos profissionais da área administrativa envolvidos com estratégia de *marketing* e planejamento estratégico. Destinado também aos profissionais que lidam com sistemas e tecnologias de informação, especialmente com *datawarehouse* e *Data Mining*. Enfim, o referido trabalho pretende atingir profissionais que trabalham com as áreas administrativa, de estratégia, de *marketing* e de sistemas de informação, além de empresas que possuam interesse em usar essa ferramenta de mineração de dados.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresenta-se a fundamentação teórica. Para o desenvolvimento do presente estudo, alguns aspectos teóricos e definições não podem ser relegados ao segundo plano, sob pena do leitor não conseguir compreender algumas definições importantes para o entendimento global da pesquisa.

Assim, se faz mister discutir a respeito de alguns conceitos como: **varejo** (e dentro deste tema, o supermercado); **gestão de sistemas ERP** (*Enterprise Resources Planning*); **banco de dados** e **data mining**.

2.1 Varejo

Las Casas e Barboza (2007) comentam sobre a existência do comércio e do varejo em épocas remotas, salientando o fato de que Roma, Alexandria e Atenas foram grandes áreas do comércio. Lembra também o fato de os gregos serem conhecidos como excepcionais comerciantes. Las Casas e Barboza (2007) destacam ainda o papel desempenhado pelo varejista e atacadista, papel este apresentado no quadro 1, sendo que o varejista adquire do atacadista, quantidade e variedade dependendo da loja e o atacadista adquire de muitos fabricantes, compram grandes variedades e quantidades.

Las Casas e Barboza (2007, p.22) explicam que, no Brasil, o varejo tal qual se apresenta hoje, teve início no século XIX a partir do advento da industrialização e do surgimento das vias e meios de transporte; e complementam o exemplo de Visconde de Mauá que foi um dos mais importantes varejistas, sendo responsável pela fundação de bancos, construção de estradas de ferro, estaleiros, indústrias e investimentos em companhias de gás. Depois de Mauá, também surgiram comerciantes e estabelecimentos como a Casa Masson (1871), as Pernambucanas em 1906, Mesbla e Anglo-Brasileira em 1913, instaladas nas principais cidades do país.

PAPEL DESENVOLVIDO PELOS VAREJISTAS E ATACADISTAS NO MERCADO	
Varejistas	Atacadistas
Adquirem do atacadista, quantidade e variedade dependendo da loja.	Adquirem de muitos fabricantes, compram grandes variedades e quantidades.
Manuseiam os produtos por meio da armazenagem, preço e exposição.	Manuseiam os produtos por meio da armazenagem, transporte e preços.
Apresentam seus produtos por meio da propaganda dirigida ao consumidor final. Pesquisa de mercado dirigida ao fabricante e ao atacadista.	Seus esforços promocionais dirigidos as varejistas apresentam as tendências para o fabricante.
Vendem para o consumidor final, oferecem horários convenientes, pessoas de vendas treinadas, estratégias de preço e de financiamento.	Vendem para o varejista com entrega eficiente, grande variedade de produtos, estratégias de financiamento e de preço.

Quadro 1: Papel desenvolvido pelos varejistas e atacadistas no mercado

Fonte: Las Casas e Barboza (2007, p.21) adaptada de Sandhusen (2000).

Las Casas (2007) ainda apresenta, (conforme quadro 2), os tipos existentes de atacadistas, definindo-os de forma geral como “[...] um tipo de intermediário que se caracteriza por não vender ao consumidor final” (LAS CASAS, 2007, p.20).

TIPOS DE ATACADISTAS	
Atacadistas de funções completas:	<ol style="list-style-type: none"> 1. Gerais: operam com ampla gama de produtos. 2. De linha: operam em determinado ramo do mercado. 3. De especialidade: revendem produtos especializados.
Atacadistas de funções limitadas:	<ol style="list-style-type: none"> 1. Transportadores (<i>truck jobber</i>). 2. Pague e leve. 3. Direto (<i>drop shipper</i>). 4. <i>Rack jobber</i>.
Agentes:	<ol style="list-style-type: none"> 1. Comissário: trabalho de corretagem; em alguns casos, assume a posse do produto. Exemplo: corretagem de café. 2. Representante comercial. 3. Corretores.

Quadro 2: Tipos de atacadistas

Fonte: Las Casas e Barboza (2007, p.21) adaptada de Simões (1984).

Lembrando que em termos históricos, o desenvolvimento mais concreto do varejo se deu logo após a Segunda Guerra, com declínio do setor atacadista tradicional, que dominava tanto o setor produtivo como distributivo. A partir daí houve um aumento da concentração dos negócios de varejo (LAS CASAS E BARBOZA, 2007).

Com relação às definições de varejo existem várias. Essencialmente, **diz respeito à comercialização de produtos e de serviços aos consumidores finais**. Nessa perspectiva, Araújo (2010) define varejo como uma atividade de venda para compradores que adquirem bens ou serviços para seu consumo. Desta forma, o que diferencia a atividade é a sua finalidade, ou seja, os compradores do varejo são também denominados de consumidores finais.

Para Estender (2005, p. 2), o varejo pode englobar qualquer atividade de venda direta de produtos e de serviços ao consumidor final, desde que seja para uso pessoal e não para negócios. Assim, qualquer empresa que esteja fornecendo seu produto e/ou serviço para o consumidor final é praticante do varejo.

Richter (1954, p. 53) define varejo como: “ [...] o processo de compra de produtos em quantidade relativamente grande dos produtores atacadistas e outros fornecedores e posterior venda em quantidades menores ao consumidor final”.

Parente (2000, p. 22) resume varejo da seguinte forma: “[...] consiste em todas as atividades que englobam o processo de venda de produtos e serviços para atender a uma necessidade pessoal do consumidor final”.

Dessa maneira, o varejista é aquele que representa o elo entre o consumidor e o atacadista e/ou produtor. Assim, pode-se definir varejo como a destinação de produtos e/ou serviços ao consumidor final em quantidades menores do que comumente é realizado pelo atacado.

Em resumo, o setor atacadista abastece o setor varejista e este é o responsável para que o produto ou serviço chegue até o consumidor final, atendendo aos seus desejos e necessidades de consumo. É neste novo cenário que entram as ferramentas de TI, dentre elas o *Data Mining*.

Esta pesquisa tem por objetivo analisar padrões de repetição de compra em um varejista de vendas ao consumidor final em um conjunto de dados, contendo itens que estão relacionados à ocorrência de outros itens (ou seja, dependents), nas operações de vendas no supermercado.

2.1.1 Supermercado

De acordo com Sebrae-ES (1999, p.1), o conceito de supermercado, enquanto comércio de mercadorias, a partir de um sistema de autosserviço, surgiu no Brasil na década de 50, sendo o “ritmo frenético dos grandes centros urbanos é o principal motivador para o aparecimento deste espaço que congrega praticidade, conforto e diversidade”.

Os supermercados são tidos comumente como varejista de alimentos. Fingerl (1996) comenta que pelos anos 50 houve a introdução do conceito de autosserviço (Peg-Pag), pela rede Pão de Açúcar; após os anos 80, com a chegada do Carrefour, consolidou-se o conceito de hipermercado. E nos anos 90 com a entrada do Wal-Mart, grupo americano, maior varejista mundial, com sua política de preços baixos e moderna tecnologia de informação sustentando eficiente logística de aquisição e distribuição.

Conforme apresentação de Fingerl (1996, p.5), os produtos alimentícios podem ser vendidos através de vários tipos e formatos de estabelecimentos comerciais. Fazendo uma retrospectiva histórica do varejo brasileiro de alimentos, os estabelecimentos comerciais comumente são classificados da seguinte maneira: autosserviço – são “os produtos alimentícios dispostos em formato *self-service* e com *check-outs* (caixas) na saída” – disponibilizados por supermercados; hipermercados e lojas de conveniência; e os tradicionais – “venda dependente de atendimento, requerendo a presença de vendedor ou balconista, com menos de três *check-outs*, caracterizadas pelo comércio de alcance local ou de vizinhança” – disponibilizados pelos armazéns, padarias e mercearias.

Na visão de Fingerl (1996), supermercado é aquele que está ao alcance de um bairro ou vizinhança, compras rotineiras, preços médios, pouca autonomia em termos de política de preços e compras, padrão arquitetônico de acordo com o perfil da área em que está instalado, predomínio de alimentos (principalmente perecíveis) no *mix*.

Quanto a definição de Hipermercado, o mesmo está localizado perto de grandes áreas onde há um intenso fluxo e fácil acesso, destinado a compras de maior volume, preços mais baixos e com mais descontos; padrão arquitetônico mais despojado; participação maior de não-alimentos no *mix*, produtos e serviços complementares na linha de acessórios para a casa, automóveis, etc.

De acordo com Lextec (2010, p.1), supermercado é uma “grande superfície comercial de venda a retalho de produtos domésticos, com uma área entre 200 e 2000 m² da qual 60% é reservada à exposição de produtos alimentares”.

Destarte, hipermercados e supermercados, cada qual com suas estratégias, através do conceito de autosserviço (Peg-Pag), trabalham na perspectiva de entender e facilitar a vida dos consumidores. No entanto, nota-se que a concorrência está cada vez mais acirrada; pois novas estratégias estão sendo montadas diariamente na intenção de angariar novos clientes e fidelizar os já existentes.

Os clientes querem ser valorizados, assim, nota-se que medidas de *marketing* estão sendo incrementadas para que estes possam ser reconhecidos como únicos. Está sendo adotado o conceito das vendas, a partir de uma perspectiva personalizada, para que este cliente se sinta atraído e tentado a voltar ao estabelecimento. A localização adequada e o *mix* de produtos diferenciados com preços atraentes já não são garantia de manutenção de seus clientes. Assim, nota-se, em resumo, uma corrida dos supermercados em montar um *mix* de serviços agregados. Nesta perspectiva o *data mining* pode ajudar a garimpar dados armazenados e antes desconhecidos, guiando o supermercado a adotar estratégias diferentes tanto na organização do *layout* como na própria exposição dos produtos, criando estratégia de vendas com foco em associação de produtos relevantes, possibilitando superar a concorrência e atrair um número maior de clientes.

2.2 Enterprise Resources Planning - ERP

Santana e Freitas (2006), definem ERP como uma espécie de pacote comercial de software de gestão empresarial que busca organizar, padronizar e integrar as informações que fazem parte da empresa.

A gestão de ERP (*Enterprise Resources Planning*) envolve várias áreas: planejamento, suporte, operação e desenvolvimento. Estas áreas devem estar atreladas aos principais atores que são os fornecedores de ERP, fornecedores gerais, consultorias, áreas usuárias, outras empresas usuárias e governança de TI, os quais atuam em dois espaços: respostas tecnológicas e necessidades do negócio respectivamente. Sem esquecer do uso das capacidades essenciais para o melhor aproveitamento da TI.

A TI surge para otimizar o tempo de respostas, sendo que o ERP ou Planejamento de Recursos Empresariais, criada a partir dos anos 90, faz parte dessa nova abordagem de sistema corporativo com vistas à informação, atuando na integração de áreas e na gestão de sua cadeia de suprimentos. O processo da gestão de TI é permeado por dois aspectos de acordo com Souza e Zwicker (2006): atendimento das necessidades do negócio e resposta tecnológica; envolvendo decisões nos mais variados prazos, levando em consideração o custo-benefício. A rápida evolução de TI e o tempo de resposta que justifiquem os gastos em tecnologia constituem alguns dos desafios da gestão de TI.

Para Souza e Zwicker (2006, p.3):

Equacionar este atendimento é um problema amplo que, entre outras coisas, passa pela definição da estratégia da empresa, pelo conhecimento da tecnologia atual e futura e pelo adequado relacionamento entre a área de TI e as áreas de negócio.

Percebe-se atualmente a grande demanda pois o ERP se mantém como prioridade de investimento em TI da maioria das empresas do mundo. No Brasil, de acordo com estudos realizados pela IDC Brasil, 23% das companhias que operam no país tem o ERP como prioridade em sua área de TI. Em 2005, 14% das empresas brasileiras investiram mais de R\$2,6 milhões na implantação de projetos de ERP e 6% investiram menos de R\$9 mil (*Computerworld*, 2006).

Silveira-Netto (2009, p.1) diz que o ERP “[...] serve para integrar todos os departamentos e funções de uma companhia em um simples sistema de computador que possa servir a todas necessidades particulares de cada uma das diferentes seções”. Comenta ainda o autor que é muito difícil construir um único programa de *software* para suprir as necessidades do departamento financeiro ou dos recursos humanos, por exemplo, mas é isso que o ERP faz. Por sua vez, Saccol et al. (2004, p.13) definem os sistemas ERP como pacotes de aplicações computacionais dando suporte à maior parte das necessidades de informação das organizações, tendo o MRP (*Manufacturing Resource Planning*) como derivado dele. Percebe-se que o ERP tem como principal diferencial a integração das informações da empresa, por meio do uso de um banco de dados único para toda a organização. Ele é composto de módulos integrados que atendem a cada área funcional ou processo, como Finanças, Produção, Custos, Vendas, RH etc.

Ainda sobre os benefícios do ERP, Perini e Raimundo (2009, p.5) salientam que o mesmo facilita o fluxo de informação entre todas as funções de uma empresa, como manufatura, logística, finanças e recursos humanos, pois, entre outros:

- Integra diversos departamentos de uma empresa;
- Possibilita a automação e armazenamento de todas as informações de negócios;
- Surgiu dos MRPs (*Material Requirement Planning*) - planejamento do uso dos insumos e a administração e etapas dos processos produtivos;
- MRP II (*Manufacturing Resource Planning*) - controla atividades como mão-de-obra e maquinário.

Além de promover agilidade na obtenção de informações e facilitar a administração sob todos os ângulos, maneiras, enfoques e objetivos diferentes de cada empresa.

A fabricação, logística, finanças e recursos humanos é dividida em módulos. Cada módulo é criado para receber e processar informações de uma área, funcionando independente dos outros módulos. Pode-se instalar um módulo do programa de cada vez ou implantar todos os módulos de uma vez - estratégia "Big Bang".

Silveira Netto (2009, p.1) afirma que essa abordagem integradora do ERP pode gerar um grande retorno financeiro à empresa se a mesma utilizar o *software* adequadamente. Por exemplo, o cliente faz um pedido. Tal pedido tem início com uma jornada de papéis que são digitados e redigitados em vários computadores da empresa, acarretando atrasos, perda de tempo e de pedidos. A digitação de dados em vários sistemas pode, da mesma forma, incorrer em erros. Igualmente, o estágio exato em que se encontra o pedido se torna difícil de mensurar, já que não há como o departamento financeiro entrar no sistema do depósito para ver se o item já foi expedido. Não raras vezes é solicitado ao cliente para que ligue ao depósito, quando o mesmo liga para qualquer outro setor querendo saber da expedição de sua mercadoria. De acordo com esse mesmo autor, há três razões principais pelas quais as empresas adotam o ERP: (1) uniformizar o processo de manufatura; (2) uniformizar as informações de RH; e (3) integrar dados financeiros.

Comenta Silveira Netto (2009, p.2) sobre essa última razão:

Como o CEO tenta entender a performance geral da companhia, ele ou ela podem encontrar diferentes versões da verdade. O financeiro tem os seus números, vendas tem outra versão, e as diferentes unidades podem, cada uma, ter a sua própria versão do quanto eles podem contribuir para a receita. O ERP cria uma única versão da verdade que não pode ser questionada porque todos estão usando o mesmo sistema.

Os principais fornecedores dos sistemas de software ERP, Baan, Totvs, JD Edwards, Microsiga, MK Group, Oracle, Peoplesoft, e SAP. Através da figura 1 pode-se observar como estava o *market share* destas empresas em 2008.

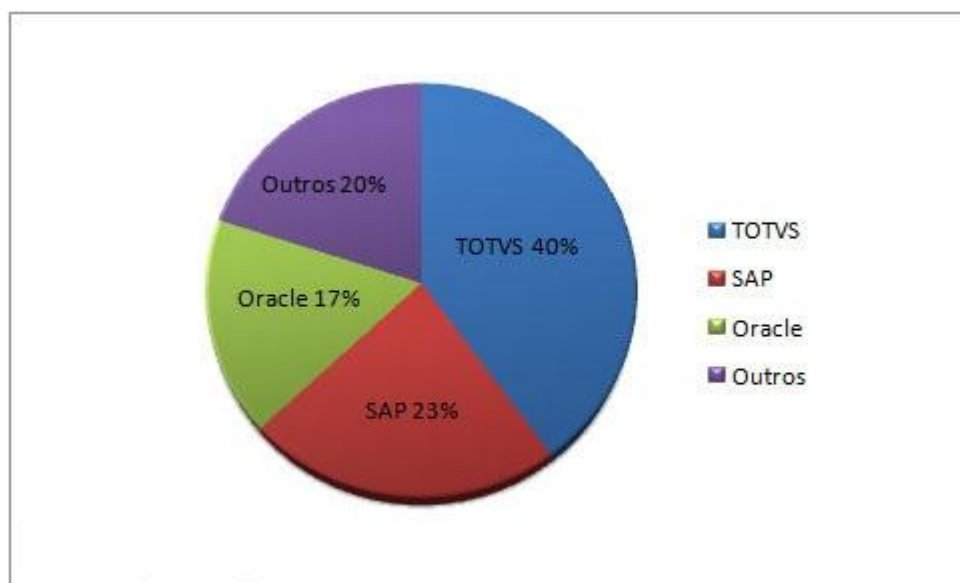


Figura 1: Market Share de ERPs no Brasil em 2008 – Fonte FGV.

Em resumo, pode-se observar que o ERP surge como ferramenta integradora de diversos departamentos da empresa. Ao construir os relacionamentos entre as mais diversas áreas de negócios tal ferramenta faz a mediação dos esforços da área de tecnologia da informação.

Entre as maiores contribuições do sistema ERP podem ser citadas: melhoria de fluxo de informações entre os departamentos; incremento do desempenho operacional; melhoria da *performance* de funcionários; melhoria do fluxo de informações na cadeia de suprimentos; redução do retrabalho; aumento da facilidade de uso dos sistemas; redução das despesas de TI e aumento do grau de terceirização de TI.

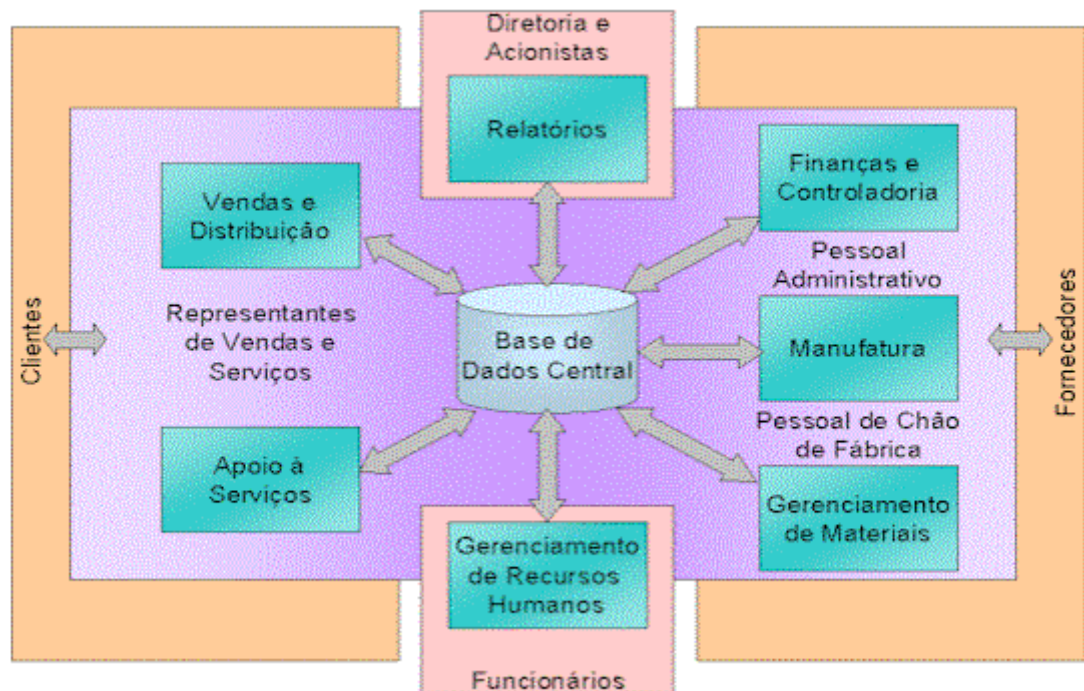


Figura 2: Estrutura típica de funcionamento de um sistema ERP
 Fonte: Silveira Netto (2009) adaptado de Davenport (1998).

Saccol et.al (2004) apresentam os resultados de uma pesquisa realizada numa amostra de 500 melhores e maiores empresas do Brasil de acordo com a Revista Exame (2000). “A pesquisa procurou avaliar o impacto da utilização dos sistemas ERP, sobre as variáveis estratégicas das organizações”. No mercado competitivo que se figura, se uma empresa não adota uma TI já utilizada pelos seus concorrentes, está em situação de desvantagem competitiva, e segundo os autores, parece ser esse o caso dos sistemas ERP. Os autores ainda lembram que algumas das empresas pesquisadas “desenvolveram seus sistemas integrados internamente, ou tiveram módulos desenvolvidos internamente, sendo utilizados paralelamente aos pacotes comerciais de ERP”, frisam ainda que “algumas delas possuem até três diferentes pacotes ERP (de diferentes fornecedores), utilizados simultaneamente” Saccol et al. (2004).

Dentre as 500 empresas analisadas, 150 não quiseram participar da pesquisa; 126 não puderam ser contatadas porque apresentavam erro de cadastro, dificultando o contato, haja visto que os contatos foram realizados por e-mail; 15 não possuíam ERP; e 209 se dispuseram a responder a pesquisa, afirmando possuir o sistema ERP implantado (Saccol et.al 2004). No decorrer das etapas da pesquisa, contudo, muitas empresas desistiram de responder à pesquisa e a amostra final

acabou ficando com 70 das 500 maiores e melhores empresas brasileiras de acordo com a Revista Exame (2000). “A maior parte delas pertence ao setor industrial (70%); seguido do setor Comércio (20%) e Serviços (10%)”.

Com relação ao bloco de clientes e fornecedores, a maioria dos respondentes, em média 87%, concordou que o ERP auxilia a empresa no suporte administrativo aos clientes, como faturamento, cobranças, gestão de estoques. “Nesse sentido, o ERP contribui estrategicamente para a melhoria das relações com os clientes da organização”. No bloco de perguntas sobre fornecedores; mais de 50% das empresas “concordam que o ERP colabora para o aumento do poder de barganha da empresa junto aos seus fornecedores” (Revista Exame 2000). O ERP também é visto por 67% dos entrevistados, aproximadamente, como recurso que auxilia na redução da incerteza de *lead time*¹ possivelmente em virtude das informações disponibilizadas em tempo real. Quase 70% dos respondentes concordaram também que a ferramenta ajuda a empresa a encontrar novas alternativas de fornecedores, auxilia nas decisões relacionadas a produzir *versus* comprar determinado insumo, bem como serve para monitorar a qualidade dos produtos e serviços recebidos dos fornecedores.

No bloco eficiência e eficácia organizacional, uma das questões levantadas foi a respeito da melhoria do processo e conteúdo das decisões a partir da ferramenta de ERP, quase 90% dos respondentes afirmam que o ERP contribui nesse sentido. Esse mesmo percentual afirma que uma das grandes vantagens do ERP é a integração organizacional; 80% concordam que essa ferramenta contribui para melhorar a avaliação dos relatórios anuais do orçamento, o que melhora a gestão financeira da empresa. “O ERP também é visto como ferramenta eficaz para o planejamento na organização: 65,7% dos respondentes concordam que ele ajuda a melhorar o planejamento estratégico” (Revista Exame 2000). Em se tratando também da gestão financeira, 52,9% concordam que o ERP ajuda a aumentar a margem de lucro da empresa, contudo 37,1% se mostraram indiferentes.

Assim, pode-se analisar que a capacidade de liderança da empresa se expressa no envolvimento da alta direção nas decisões sobre o ERP. A visão

¹ *Lead time* é o tempo que decorre entre a liberação de uma ordem (de compra ou produção) e o momento a partir do qual o material referente à ordem está pronto e disponível para uso (SACCOL *et. al.* 2004).

sistêmica de negócios parece ser incitada pela tecnologia e é parte integrante do cotidiano da gestão ERP nas organizações.

2.3 Banco de Dados

Pode-se dizer que os dados são informações brutas que sozinhas não possuem relevância, a menos que trabalhadas. Ao se mencionar banco de dados, pode-se frisar que se trata de um conjunto de dados armazenados para um objetivo específico. Quando organizados, os mesmos produzem a informação. Assim, define Dias (2008, p. 1) “Um banco de dados é uma coleção de dados relacionados. Definimos dados como fatos conhecidos que podem ser registrados e que possuem significado implícito. Por exemplo: Agenda telefônica e de endereços.”

Segundo Dias (2008), um Banco de Dados é modelado com dados com um objetivo já especificado, pois já possui um grupo provável de usuários e algumas aplicações preconcebidas, nas quais esses usuários estão interessados. Sendo que este banco de dados possui alguma fonte, da qual os dados são derivados, algum grau de interação com eventos do mundo real e um público que está ativamente interessado no conteúdo do banco de dados, que pode ser gerado e mantido manualmente ou pode ser informatizado. O catálogo de fichas de uma biblioteca é um exemplo de banco de dados que pode ser criado e mantido manualmente. Um banco de dados informatizado pode ser criado e mantido por um grupo de programas de aplicações, escritos especificamente para aquela tarefa ou por um sistema de gerenciamento de banco de dados. Portanto, um banco de dados pode ser definido como um conjunto de dados organizados de forma lógica e coerente, os quais por sua vez, possuem significado inerente.

Wilson (2010, p.5) define SGBD² como um *software* que apresenta recursos específicos de forma a facilitar a manipulação de informações dos bancos de dados e o desenvolvimento dos programas aplicativos. O mesmo autor comenta que existem alguns modelos de banco de dados, entre eles estão o modelo **conceitual** que não contém detalhes sobre a representação em meio físico das informações; o modelo **lógico** que apresenta a descrição de como as informações estão organizadas internamente, visão do usuário do SGBD; e o modelo **físico** que

² SGBD - Sistema de gerenciamento de banco de dados.

descreve os dados no nível mais baixo (interno); trata dos aspectos de implementação do SGBD (WILSON, 2010, p.5).

O modelo lógico especificamente divide-se em três outros modelos: o modelo de rede (o qual é representado por um conjunto de registros, cujas relações são registradas através de ponteiros); o modelo hierárquico (o qual é similar ao modelo de rede, diferenciando-se pela representação gráfica, sendo seus registros organizados em árvores); e o modelo relacional (o qual faz uso de um conjunto de tabelas para a representação de dados, as quais são compostas por linhas e colunas) (Wilson 2010). A Figura 3 ilustra os três modelos lógicos.

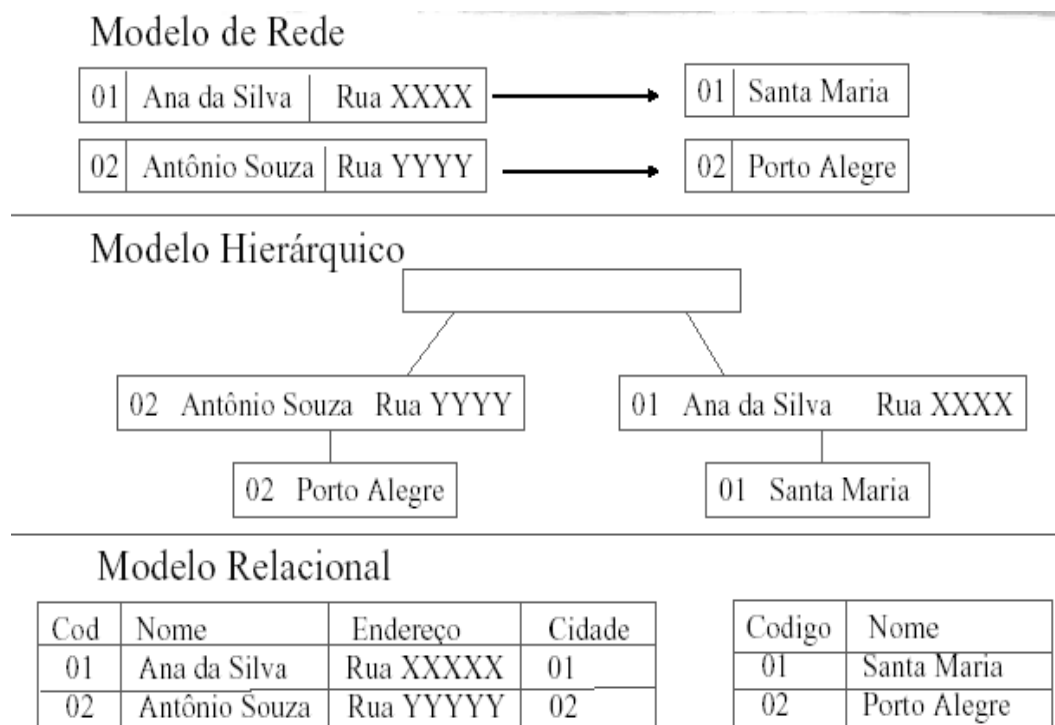


Figura 3: Modelos lógicos de banco de dados: Rede, Hierárquico e Relacional.
Fonte: Wilson (2010).

É nesta perspectiva que nasce o *Data Warehouse* (DW) e, por conseguinte, o *Data Mining*. A função do *Data Warehouse* é tornar as informações corporativas mais acessíveis para entendimento e uso das áreas estratégicas. Observa-se que o projeto de *Data Warehouse* não é uma tarefa fácil, pois envolve um conjunto de conceitos e tecnologias. Destacando-se que o sucesso de um projeto de *Data Warehouse* está relacionado com o entendimento e domínio destes conceitos e tecnologias. A causa principal que resulta em falha e insucesso de um projeto de *Data Warehouse* está relacionada à ausência de uma metodologia abrangente

capaz de fornecer uma visão geral do processo envolvendo estes conceitos e tecnologias.

Ao se criar um ambiente de *Data Warehouse* seu aspecto fundamental diz respeito à separação dos dados do ambiente operacional de fontes diversas para o ambiente de *Data Warehouse*, conforme demonstra a figura 4.

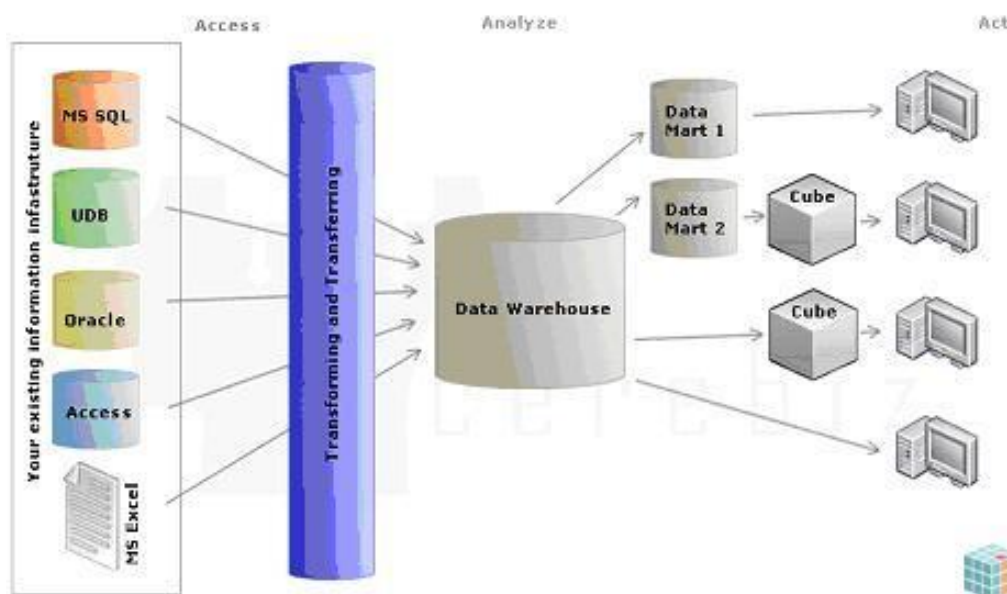


Figura 4: Arquitetura Data Warehouse
Fonte: Serain (2007, p.3).

Como analisado na ilustração, todo ambiente de *Data Warehouse*, ou depósito de dados como é chamado também no Brasil, depende de *Data Marts*. Após os dados serem coletados dos sistemas transacionais (OLTP), eles possuem rotinas de ETL (Extração, Transformação e Carga dos Dados) que trazem as informações de ambientes externos fazendo posteriormente tratamentos, agregações e sumarizações.

No supermercado em estudo, o software utilizado é o Conexo, composto dos módulos: cadastros, faturamento, outros movimentos, inventário fiscal, cotação, pedidos, devoluções, pesquisas, industrializados, preços, custos, estoques, mata-burro, vasilhame, financeiro, integração, apoio, pdv, fechamentos, comunicação, relatórios e gerência. As informações são armazenadas seguindo o modelo lógico relacional, contudo, não existe a preocupação do supermercado de como os dados estão sendo estruturados internamente. Os profissionais de TI atuam mais no suporte do sistema e os profissionais que alimentam o sistema tem o foco no

cadastro de produtos e serviços, controle de estoque, sem a percepção que todos estes dados podem trazer informações relevantes para melhorar o desempenho da organização. Ficou constatado que alguns produtos são cadastrados sobrepondo produtos que foram descontinuados. Por consequência, quando existe problemas fiscais ou mesmo quando da avaliação com o *Data Mining* foco desta pesquisa, as informações podem ficar prejudicadas.

Construir um projeto de *Data Warehouse* é dispendioso e demorado, e, pensando nisso, as organizações têm procurado começar sua construção pelos departamentos, fazendo o processo inverso, ou seja, ao invés de construir um DW e a partir dele os *Data Marts* (DM), começam pelo DM e esses fazem a alimentação no DW.

Bonomo (2009, p.3) explica que os *Data Marts* são bancos de dados modelados multidimensionalmente, orientados aos departamentos de uma empresa, dando como exemplo, uma área comercial, financeira, marketing e logística constituindo vários *Data Marts* (DM's) cada uma. Ao final de dois anos, os departamentos passaram a alimentar um DW que, depois de implementado, passou a fazer o caminho inverso, que é de alimentar os DM's. Com essa fundamentação, podemos definir que os DM's podem surgir de duas maneiras. A primeira é *top-down* e a segunda é a *botton-up*.

Bonomo (2009, p.3) define como *Top-down* quando a empresa cria um DW e depois parte para a segmentação, ou seja, divide o DW em áreas menores gerando assim pequenos bancos orientados por assuntos aos departamentos. *Botton-up* é uma situação inversa, ou seja, a empresa, prefere primeiro criar um banco de dados para somente uma área. Com isso os custos são bem inferiores de um projeto de DW completo. A partir da visualização dos primeiros resultados parte para outra área e assim sucessivamente até resultar em um Data Warehouse. A infra-estrutura de software e hardware é parecida, porém a arquitetura dos dados pode ser bem diferente. No DW pode-se ter o modelo de dados híbrido ou pode ser completamente relacional, já no DM o modelo é totalmente dimensional. A complexidade de tratamento das informações no DW é maior, porque tem que se olhar toda a estrutura da empresa, já no DM fica mais fácil, pois estamos olhando apenas uma parte dela. Assim, é função da empresa avaliar a sua demanda e optar pela melhor solução (BONOMO, 2009, p.3).

Bonomo (2009, p.1) ainda caracteriza os projetos de DW e DM e cita as tecnologias usadas pelos mesmos em suas diferentes fases, sendo:

a) Data Warehouse (DW):

- Dados integrados e organizados por assunto
- Histórico de informações da empresa
- Esquema estrela (fatos e dimensões)
- Não-Volátil
- Metadados
- Grande volume de dados

b) Data Mart (DM)

- Dados agregados
- Não possui necessariamente um histórico das informações
- Esquema estrela (fatos e dimensões)
- Pode ser reconstruído a partir dos dados existentes no DW
- Pequeno volume de dados
- Otimizado para consultas pontuais

c) Tecnologias Utilizadas

- Banco de dados (Relacional ou Multi-dimensional)
- ETL (*Extract, Transform and Load*)
- OLAP (*On Line Analytical Processing*)
- *Data Mining*

Assunção (2001, p.1) ressalta que se o DW não tiver conexão “[...] com ferramentas de análise/extração é simplesmente um banco de dados sem utilidade. Por outro lado, atinge sua plenitude quando é modelado de tal forma a permitir uma ampla compatibilidade com a maioria das ferramentas existentes”. O *Data Mining* assim é uma ferramenta utilizada no DW e será esmiuçado a seguir (Bonomo 2009).

Cenários competitivos como os atuais exigem suporte às estratégias para que se possa formar uma base sólida nos negócios do mundo corporativo. Sob essa perspectiva, o supermercado tem utilizado o *software* mais para controladoria, não

se atendo a importância de um Data Warehouse ou mesmo Data Mart, sem visualizar a importância estratégica que poderia ser o uso das informações estratégicas extraídas do enorme banco de dados que possui.

2.4 Data Mining

Com o advento da globalização, informações de qualidade são cruciais para o sucesso organizacional. Pode-se dizer assim que as tecnologias de informação, as ferramentas e as técnicas utilizadas para melhor obter, tratar, apresentar e disponibilizar informações que são primordiais para o aumento da competitividade organizacional.

A crescente necessidade de discernir a informação útil de grandes quantidades de dados armazenados em bancos de dados gigantes, que até a década de 1980 era realizada através da análise estatística tradicional, ganham um novo aliado: *Data Mining*.

Para apresentação do tema será utilizada a definição de Lamboia e Pereira (2005, p.9), os quais destacam o *Data Mining* como o processo de seleção, exploração de dados e criação de modelos que utilizam um grande volume de armazenamento de dados para identificar padrões previamente desconhecidos. Consiste na construção de modelos computacionais para a descoberta automática de novos fatos e relacionamentos entre dados, produzindo novos conhecimentos. Esses conhecimentos são utilizados para informar melhor os tomadores de decisão antes de eles agirem, podendo-se criar um modelo do mundo real com base nos dados coletados de várias fontes, incluindo transações corporativas, dados históricos e demográficos de clientes, e mesmo de fontes externas, como empresas de análise de crédito. Depois, pode utilizar esse modelo para produzir padrões a partir de informações, os quais podem dar suporte às tomadas de decisão e prever novas oportunidades de negócios.

O *Data Mining* assim está presente em vários setores e atividades empresariais, no sentido de detectar fraudes, identificar estratégias mais lucrativas, avaliar a eficácia de promoções, descobrir padrões comportamentais nos clientes e otimizar, por exemplo, campanhas de *marketing*. Côrtes (2004, p. 122) diz que os “bancos de dados da empresa são vasculhados e analisados em busca de informações que, embora sempre estivessem lá, não eram conhecidas. Isso é feito

utilizando *softwares* específicos [...]”. Tal afirmação vem reforçar a importância da ferramenta de mineração de dados no sentido de transformar dados em informações relevantes.

Por mineração de dados, entende-se o processo através do qual se extrai informações desconhecidas de bases de dados, utilizando-as para a tomada de decisões crítica dos negócios (GONÇALVES, 1999, p.5). A esse respeito: “os softwares de mineração de dados são capazes de peneirar grandes volumes de dados para encontrar ‘pepitas’ de informação as quais produzem ‘ouro’ em forma de vantagem competitiva” (GONÇALVES, 1999, p.5). Dessa maneira, as informações encontradas a partir do *Data Mining* podem gerar vantagem competitiva para as organizações, haja visto que é possível traçar, dentre muitas outras coisas, o perfil da clientela podendo assim, antecipar-se à concorrência.

O surgimento de uma economia globalizada levou as organizações a transformarem seus processos nos negócios, tornando-se cada vez mais dependentes da informação. Isto demanda uma infraestrutura tecnológica que permite um gerenciamento mais eficiente desta grande quantidade de dados. Através de suas diversas técnicas, o *Data Mining* busca explorar e identificar através de padrões desconhecidos, conhecimentos que possam dar suporte ao gestor na tomada de decisões e ainda possibilitar a garimpagem de novas oportunidades de negócios. Na figura 5 verifica-se o modelo do processamento de *Data Mining*.

Lamboia e Pereira (2005) apresentam os setores e atividades de negócio que usam o *Data Mining*, bem como, de que forma essas empresas usam a ferramenta:

- Empresas de telecomunicações, seguros, cartões de crédito e do mercado de ações utilizam o *Data Mining* para detectar fraudes, otimizar campanhas de marketing e identificar as estratégias mais lucrativas.
- Área médica utiliza o *Data Mining* para prever a eficácia de procedimentos cirúrgicos, exames médicos e medicações.
- O setor de varejo utiliza o *Data Mining* para avaliar a eficácia de promoções e eventos especiais, e prever que ofertas são mais apropriadas para os diferentes consumidores.

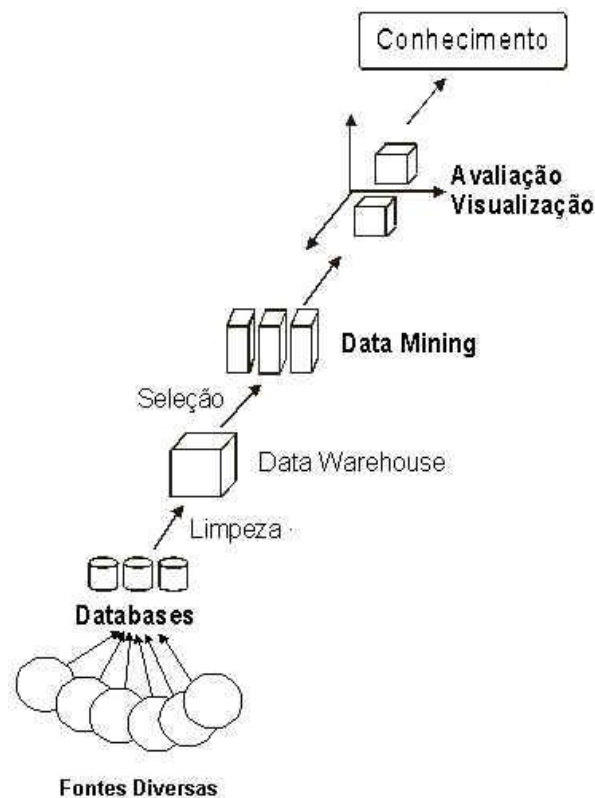


Figura 5: Processamento *Data Mining*
 Fonte: Navega (2000, p.2)

No supermercado em estudo, até então, não havia sido utilizado nenhum estudo no banco de dados ou especificamente de *Data Mining* para auxiliar na descoberta de informações que proporcionassem verificar a eficácia de promoções e que pudesse ainda prever ofertas mais adequadas para consumidores diferentes. As estratégias nas ofertas e promoções acontecem pela negociação entre comprador e fornecedor, queima de estoque, data de validade dos produtos, dentre outros. Tem-se como expectativa que com a descoberta de associações relevantes, o supermercado possa melhorar suas campanhas de marketing, fidelizando clientes e identificando estratégias mais lucrativas.

A partir da mineração de dados ainda é possível traçar o perfil do consumidor, montar estratégias para alavancar vendas, analisar formas para atrair e fidelizar clientes, dentre outros. Ademais, não há alguma área específica para a aplicação do *Data Mining*. Dentre as que mais têm explorado esta ferramenta, pode-se citar: marketing, finanças, vendas, manufatura, saúde e energia.

O marketing de relacionamentos é uma das ramificações do marketing bastante comentada nos últimos tempos. Em termos técnicos é chamada de *CRM* –

Customer Relationship Management. De acordo com Cabrino (2002, p.1) é: “uma nova face dos negócios, um conjunto de oportunidades estratégicas, no qual o consumidor torna-se o centro de toda atenção da organização em tempo integral”. Claro (2006, p.3), professor pesquisador do Ibmec São Paulo, destaca que os relacionamentos permitem que empresas explorem ao máximo os seus recursos como objetivo de obter vantagens competitivas. Os mercados estão passando por uma situação em que as empresas necessitam oferecer valor aos clientes, precisam se diferenciar frente a uma concorrência que utiliza cada vez mais de modelos de marketing agressivos. É crítico para uma empresa ter um desenvolvimento de produtos impecável, definir preços competitivos, desenvolver uma campanha promocional eficaz e tornar seus produtos e serviços disponíveis de forma que o cliente tenha um conveniente acesso a eles. Acima de tudo, o cliente precisa ser fidelizado. O tema do relacionamento é fascinante, pois permite que empresas, competindo em mercados onde a concorrência atingiu níveis elevados, criem valor para o cliente.

Através do marketing de relacionamento, as empresas projetam e lançam ações direcionadas aos clientes, procurando fidelizá-los e não somente isso, mas espera-se que tais clientes utilizem e indiquem os produtos e/ou serviços para terceiros, reconhecendo o valor e o esforço daquela empresa que se preocupa e atende as solicitações de sua clientela.

Para Claro (2006, p.4), a orientação de marketing, que busca produzir o que os clientes querem e precisam, é baseada em três pilares: “orientação para o cliente, coordenação e integração de todas as atividades de marketing, e foco na criação de valor para os stakeholders”. Cabrino (2002, p.1) salienta, entretanto, um outro lado da situação, que todo relacionamento requer um conhecimento, e desta forma os consumidores acabam perdendo sua privacidade, pois preenchem constantemente intermináveis formulários.

Segundo Cabrino (2002), explorar o conhecimento é indispensável para a realização do marketing de relacionamento, mas saber coletar e trabalhar os dados é bem diferente e se não for realizada de forma profissional, poderá acarretar em inúmeros prejuízos para as organizações que possuem este tipo de prática.

Inegavelmente, no entanto, o relacionamento está presente no sistema de comercialização fazendo parte de todo o processo de venda. Conhecendo os hábitos e os costumes de seu cliente, a organização pode antever as ações de seus

clientes e criar estratégias. Contudo, este tipo de relacionamento não é praticado em grande parte de segmentos, fazendo com que as organizações percam bastante espaço no mercado. Assim, para Cabrino (2002), o consumidor pode vir a conhecer um outro estabelecimento, que possua uma qualidade nos produtos semelhantes a este, mas que pratique um serviço diferenciado, vindo o mesmo a migrar, deixando de lado a antiga, possivelmente não mais realizando seus pedidos lá.

Para Claro (2006, p.6), o marketing de relacionamento está no topo da agenda de ações de gestores e pesquisadores por diversos fatores. O primeiro citado é “a grande propagação de sistemas de informação que integra os negócios em uma cadeia de valor e permite rapidamente se ter acesso a informações valiosas”. De acordo com Claro (2006), “muitas empresas, as equipes de informática vêm refinando softwares e processos a fim de aferir os custos e despesas e permitir uma estimativa do valor entregue aos clientes”. Destarte, as empresas conseguem inclusive estimar quanto custa perder um cliente. Assim, é feita a análise de quanto se gasta para atrair novos clientes para substituir aqueles que estão saindo e quanto à empresa gasta para manter os antigos. A esse respeito diz o autor:

A Consultoria Mckinsey realizou recentemente uma pesquisa demonstrando que compras repetidas (clientes fiéis) geram mais que o dobro de lucratividade quando comparado com novos clientes. A conclusão dos consultores é de que clientes devem ser tratados como ativos ou investimentos a serem criteriosamente administrado. Isto nos leva a crer que o sistema de informação é um fator que leva empresas a adotarem uma orientação para o marketing de relacionamentos, pois os sistemas permitem as empresas mensurarem os ganhos e perdas resultante de cada cliente e permitem também a troca de informações valiosas (CLARO, 2006, p.6).

O segundo fator apontado por Claro (2006, p.6) é com relação à percepção por parte dos profissionais de marketing da importância de atentar para o lado *soft* da interação com os clientes. Claro comenta que muitas organizações perceberam que não basta ter excelente produtos, processos impecáveis ou o melhor preço se o tratamento com a clientela não for dos melhores, ou seja, se o atendimento não for adequado. Com o avanço em serviços, os profissionais responsáveis por relacionamentos puderam desenvolver uma visão holística de como o marketing contribui para se atingir a satisfação dos clientes na sua plenitude do termo (Claro 2006).

É importante salientar que cada consumidor é único e possui necessidades diferentes, mas existem alguns fatores que influenciam a compra. Assim, é necessário para a organização compreender o mercado atual e, sobretudo, descobrir quais são estes elementos ensejadores da compra ou da contratação de um serviço. Um fator, contudo, que deve ser levado em consideração são as diferenças culturais para a implementação de qualquer estratégia, assim como, deve-se analisar de forma detalhada a importância dessas diferenças para ações de *marketing*, pois nem sempre uma medida adotada aqui no Brasil terá sucesso fora do país ou vice-versa; tudo vai depender da cultura local.

A competitividade é outro fator que as empresas precisam ter a capacidade de competir em diferentes níveis, local e global (Claro, 2006). O autor destaca que há alguns anos, as empresas que oferecessem um produto de alta qualidade garantiriam uma vantagem sobre a concorrência. Mas atualmente já não é mais assim, pois a maioria dos fabricantes de produtos possuem padrões alto de tecnologia e avanços em processos que raramente um produto apresenta mau funcionamento. Além do que, empresas respeitam os termos de garantia e quando existe a necessidade produtos são substituídos. No setor de serviços, podem-se notar os mesmos avanços e altos padrões de qualidade. Neste contexto, empresas que competem nestes mercados começam a perceber o relacionamento como um diferencial dificilmente copiado pela concorrência. Desenvolver relacionamento para estas empresas significa competir em um nível em que poucos estão se inserindo.

Dessa maneira, é possível concluir que qualidade nos produtos atualmente não é mais diferencial e sim é obrigação, ou seja, é o mínimo exigido pelos consumidores e neste cenário, as organizações devem se movimentar para oferecer diferenciais, como a qualidade nos serviços por exemplo. A globalização tem intensificado as atividades do *marketing*, já que é necessário alcançar o cliente onde quer que ele esteja. Neste bojo entra o papel primordial da logística, podendo agregar valor aos produtos e serviços, criando assim vantagem competitiva.

Claro (2006, p.7) resume os fatores que estimulam a ênfase no marketing de relacionamentos. São eles: propagação dos sistemas de informação; expansão do setor de serviços; mudança na natureza da competição; satisfação do cliente no longo termo; e formação de rede de negócios. Lamboia e Pereira (2005) citam ainda outras áreas onde a mineração de dados é utilizada, como finanças, manufatura, saúde, energia e vendas, também variável analisada no presente estudo.

A respeito das vendas, Lamboia e Pereira (2005) comentam que no setor de vendas a aplicação provavelmente de maior interesse seja a de identificar produtos que possam ser colocados em uma mesma cesta ou pacote. Isto envolve a “garimpagem” por associação entre produtos, que pode revelar afinidades ou aversões nunca imaginadas e como consequência, sugerir estratégias para maximizar o lucro. A descoberta de que dois ou mais produtos têm suas vendas fortemente associadas pode, por exemplo, ser usada no sentido de não colocá-los em promoção ao mesmo tempo, organizando de forma que o cliente o encontre rápido.

Outra aplicação, relatada por Lamboia e Pereira (2005), é a busca por associações que envolvam uma componente temporal, isto é, aquela que revela, por exemplo, que a compra de um produto hoje induz, com alta probabilidade, a compra de outro algum tempo depois. Tal descoberta pode sugerir uma estratégia de venda que busque efetivar a venda de ambos logo na primeira oportunidade.

Assim, vale ressaltar que as técnicas de associação utilizadas na mineração de dados, são aliadas importantíssimas para a formação de promoções e de vendas associadas de produtos. É possível através desta técnica, identificar um ou mais produtos, que podem incentivar de alguma forma a vendagem de outro produto. Em se tratando de supermercado, quando uma rede descobre tais produtos, podem agrupá-los de maneira que fiquem visíveis aos consumidores, incentivando assim, a aquisição de ambos os produtos. É viável também a partir da técnica de associação montar combos com produtos afins, sinalizando promoções, alavancando vendas e consequentemente otimizando o lucro.

A área de finanças, por sua vez pode utilizar de diversas técnicas de mineração de dados, como a associação, a predição, a classificação, a agregação. A mineração de dados é usada nesta área para que se possa analisar crédito, avaliar potenciais fraudes com cartão de crédito, avaliar riscos, previsão de falências, dentre outras. Lamboia e Pereira (2005, p.19) salientam que nesta área as transações podem envolver volumes financeiros efetivamente elevados, os riscos e a concorrência também são altos e neste cenário, qualquer oportunidade de ganho é considerada e pode representar um diferencial, por menor que seja à margem de ganho.

Em termos de diagnóstico o emprego de *Data Mining* visa o desenvolvimento de sistemas capazes de realizar diagnósticos automaticamente, com base em dados

obtidos em exames laboratoriais. Outras áreas que utilizam da mineração de dados são especificadas por Lamboia e Pereira (2005, p.19) como:

- a) Manufatura: a complexidade dos modernos parques de produção e a pressão pela eficiência e pela qualidade tem possibilitado o uso de *Data Mining* e de automação em diversas áreas. CAD/CAM (*Computer Aided Design/Computer Aided Manufacturing*) e robôs são algumas das aplicações de maior demanda na área.
- b) Saúde: São basicamente duas frentes distintas de trabalho nesta área: administração e diagnóstico. Na administração os sistemas lidam com os serviços oferecidos aos pacientes, com os seguros, com as ações potencialmente fraudulentas, etc.

Sistemas de diagnóstico apresentam a vantagem de atender rapidamente grandes volumes de pacientes, o que facilita a ação pública em epidemias ou campanhas sanitárias.

- c) Energia: Previsão de consumo e previsão de falhas em sistemas de transmissão ou de distribuição são as duas aplicações mais comuns, embora muitas outras tenham sido pesquisadas e difundidas na literatura. A tecnologia de maior emprego atualmente em *Data Mining* é a de redes neurais.

A figura 6 mostra as etapas do processo de mineração de dados:



Figura 6 – Etapas do processo de Mineração de Dados
 Fonte: Rezende, Pugliesi, Melanda e Paula (2003).

De acordo com Rezende, Pugliesi, Melanda e Paula (2003), explica-se neste tópico as etapas do processo do *Data Mining*:

Identificação do Problema – Nessa fase são identificados estudos do domínio da aplicação, a definição de objetivos e de metas a serem alcançados no processo de mineração de dados. Também é realizada a identificação e seleção dos conjuntos de dados a serem utilizados durante o processo.

Pré-Processamento – Normalmente os dados selecionados para o processo de mineração de dados não estão no formato adequado para a extração do conhecimento. Decorrentes de erros de digitação, erros gerados na leitura de dados por sensores e outros dispositivos, nesse sentido é necessário adequar os dados antes da aplicação de extração de padrões.

Extração de Padrões – A etapa de extração de padrões é guiada pelo cumprimento dos objetivos definidos na identificação do problema, com a execução de um ou mais algoritmos para a extração de conhecimento.

Pós-Processamento – Nesta etapa de pós-processamento de mineração de dados, o conhecimento extraído pode ser simplificado, avaliado ou simplesmente documentado para o gestor e ou departamento interessado nas informações. A de se verificar se as informações atendem aos interesses do usuário ou se não atendem aos objetivos preestabelecidos, pode ser necessária a realização de ajustes ou ainda todo o processo de extração.

Utilização do conhecimento – O objetivo maior do processo de extração de conhecimento é a utilização do conhecimento adquirido. Podendo ser incluído a um sistema inteligente, utilizado diretamente pelo usuário final para apoio a algum processo de tomada de decisão.

Na mineração de dados, tem-se tarefas e técnicas as quais precisam ser compreendidas para melhor interpretação deste trabalho. A **tarefa** consiste na especificação do que estamos querendo buscar nos dados, que tipo de regularidades ou categoria de padrões temos interesse em encontrar, ou que tipo de padrões poderiam surpreender (por exemplo, um gasto exagerado de um cliente de cartão de crédito fora dos padrões usuais de seus gastos). A **técnica** de mineração consiste na especificação de métodos que nos garantam como descobrir os padrões que nos interessam (i.e. regras de associação). Dentre as principais técnicas utilizadas em mineração de dados, temos técnicas estatísticas, técnicas de aprendizado de máquina e técnicas baseadas em crescimento-poda-validação.

Para que se entenda melhor sobre as ferramentas de mineração de dados, Gonçalves (1999, p.8) aponta as diferenças entre dados, informação e conhecimento, conceituando tais conceitos.

Basicamente, se atribuímos algum significado especial a um dado, este se transforma em informação. Se os especialistas no domínio do problema elaboram uma norma (regra), a interpretação do confronto entre esta informação e essa regra constitui um conhecimento [...]Para Freitas et al. (1997) a informação é considerada como um dado dotado de relevância e propósito, para cuja conversão é necessário conhecimento.

É possível dizer que dados, do ponto de vista organizacional, são informações brutas sobre um evento que sozinhas não possuem relevância, mas são imprescindíveis para a formação de uma informação. Esta por sua vez, é a junção dos dados e possuem emitente e receptor, já possui relevância e quando bem analisada pode gerar o conhecimento. Sendo assim, o conhecimento é derivado da informação, e esta por sua vez, dos dados.

Importante também para a tese em questão é comentar sobre a tomada de decisões e os níveis organizacionais. A esse respeito Gonçalves (1999, p.10) lembra que as decisões são tomadas em todos os níveis organizacionais, porém elas são diferenciadas. Em se tratando do nível operacional as decisões a serem tomadas são, em grande maioria programadas ou programáveis, ou seja, por tratarem de problemas rotineiros elas podem ser formalizadas em manuais, estatutos, etc. Quanto ao nível estratégico, onde encontram-se os executivos mais preocupados com aspectos externos à organização (tais como: concorrência, governo, mercado, etc.), as decisões são na maior parte das vezes não programadas, ou seja, a decisão deverá partir da análise da pessoa responsável pela mesma, não havendo regras e nem documentos que formalizam a ação que deve ser tomada em tal situação. Neste ponto é que encontra-se atualmente o maior potencial de utilização das tecnologias de informação, pois, devido à alta competitividade empresarial aquele que obtiver informações relevantes em tempo hábil e formato adequado estará à frente dos seus concorrentes.

Assim, é importante frisar que a decisão nunca é apenas uma decisão, ela pode desencadear uma série de eventos e, por esse motivo, o tomador de decisão, independente do nível organizacional ou hierárquico, deve ter subsídios para agir em consonância com a estratégia da empresa, não apenas agindo intuitivamente, mas com o auxílio de ferramentas que possam auxiliá-lo nesse momento decisivo.

Ainda a esse respeito, Gonçalves (1999, p.14) comenta que “a medida em que o nível de decisão se desloca para os níveis superiores da pirâmide, aumenta a incerteza e o risco. As decisões no nível estratégico são geralmente tomadas em uma situação de incerteza e risco”. Neste cenário o uso de TI se faz presente.

Prospectar conhecimento em uma base de dados significa identificar e reconhecer padrões, a partir da automação. Nesse cenário pode-se citar a mineração de dados como um passo no processo de prospecção, que segundo Gonçalves (1999) consiste na utilização de algoritmos que possam produzir uma enumeração particular de padrões.

A mineração de dados, também conhecida como *Descoberta de Conhecimento em Bases de Dados*, é a “extração não trivial da informação importante, implícita, previamente desconhecida, de dados” (Gonçalves 1999). Assim, ela usa de técnicas estatísticas, do aprendizado da máquina e da

visualização para que os conhecimentos sejam descobertos e apresentados de uma forma mais facilmente compreensível às pessoas.

Mineração de dados tem sido descrita como a interseção entre a inteligência artificial, aprendizagem da máquina e tecnologias de bancos de dados. Muitas vezes a meta é construir automaticamente um modelo de software que prediga um valor de saída dado um conjunto de valores de entrada. Uma variedade de técnicas podem ser usadas, e cada uma tem sua estrutura própria (GONÇALVES, 1999, p.14).

Nesta perspectiva, entende-se a mineração de dados como um conjunto de técnicas com estruturas peculiares que predizem uma saída a partir de um conjunto de entradas. Tais saídas são avaliadas como informações valiosíssimas que podem representar um diferencial competitivo para o negócio.

No que diz respeito ao varejo, pode-se dizer que para os pequenos varejistas é mais fácil fidelizar sua clientela, dado o volume de pessoas que passam pelo estabelecimento. Situação diferente ocorre com os grandes comércios. Assim, surgem questionamentos sobre de que forma grandes empresas podem atender as necessidades e lembrar das preferências dos clientes. Gonçalves (1999, p.15) faz a seguinte pergunta: *“O que pode substituir a intuição da pessoa que conhece os clientes por nome, fisionomia e voz, e lembra dos hábitos e preferências?”*.

A resposta pode vir a partir do uso inteligente da tecnologia da informação. Usando-se adequadamente as ferramentas de TI é possível mesmo para a maior empresa, ficar próxima dos seus clientes. Algumas tarefas desempenhadas pela mineração de dados: classificação, estimativa, previsão, agrupamento por afinidade ou análise de cesta de supermercado, segmentação ou *clustering* e descrição.

Anualmente, as empresas acumulam grandes quantidades de informações, entretanto, apesar do volume de dados, a maioria das companhias não consegue aproveitar o que está armazenado nos arquivos (GUIMARAES, 1999).

Nesta perspectiva a tecnologia exerce papel de protagonista, pois quando não é usada com sabedoria, pode se tornar uma das grandes responsáveis por este acúmulo de dados, excedendo a habilidade técnica e a capacidade humana para sua interpretação em muitos casos; em contrapartida quando usada de maneira apropriada é ensejadora do tratamento destes dados, gerando informações. É neste contexto que entra a importância da tecnologia da informação.

Lamboia e Pereira (2005, p.20) citam exemplos e empresas que utilizam *Data Mining* e seus respectivos benefícios entre eles se destacam:

- A rede americana Wall-Mart, pioneira no uso de *Data Mining*, descobriu ao explorar seus números que 60% das mães que compram boneca Barbie, levam também uma barra de chocolate.
- O banco Itaú conseguiu aumentar sua taxa de retorno nas malas diretas para 30%, reduzindo a conta do correio a um quinto. Isto foi feito armazenando e analisando a movimentação financeira de seus 3 milhões de correntistas nos últimos 18 meses.
- A Sprint, um dos líderes no mercado americano de telefone de longa distância, desenvolveu, com a análise de *Data Mining* em seu *Data Warehouse*, um método capaz de prever com 61% de segurança se um consumidor trocava de companhia telefônica dentro de um período de dois meses. Com um marketing agressivo, conseguiu evitar a deserção de 120.000 clientes e uma perda de 35 milhões de dólares em faturamento.
- O governo de Massachusetts, nos Estados Unidos, compilava informações financeiras imprimindo telas e mais telas nos terminais de grande porte. Só com papel foram economizados U\$ 250.000,00 por ano. Além de permitir que o orçamento anual fosse assinado e analisado, antes do início do ano fiscal.
- O SERPRO no Brasil, implantando o seu *Data Warehouse* e *Data Mining*, já consegue hoje cruzar e analisar informações em cinco minutos, o que antes demandavam quinze dias de trabalho (LAMBOIA e PEREIRA, 2005, p.20).

No supermercado pesquisado a utilização do software tem foco nos controles de cadastro de produtos, clientes, estoques e não há preocupação com os dados armazenados no sentido de transformação de informações diferenciadas. Assim, fica a expectativa que com o uso do *Data Mining* o supermercado consiga descobrir itens com suporte e confiança relevantes e de posse deste conhecimento consiga traçar novas estratégias de vendas, maximizando tempo e recursos, se aproximando cada vez mais o consumidor.

Importante salientar, contudo, que transformar dados em informações relevantes nem sempre é uma iniciativa rápida, precisa e barata e que de nada vale investir em informações valiosas se não forem usadas de forma efetiva.

Gonçalves (1999) ressalta que os usuários vêm usando diferentes ferramentas entre elas: *query*, servidores, sistemas de informações gerenciais, entre outros para examinar os dados que possuem, mas a maioria dos analistas tem reconhecido que existem padrões, relacionamentos e regras por trás dos dados que não podem ser encontrados utilizando os métodos tradicionais.

Na figura 7, os dados transformados em informações são entregues de forma desordenada, forçando o usuário a criar hipótese que será verificada pelo sistema quanto à sua veracidade ou falsidade. Já, na figura 8, os resultados aparecem ordenados, na mineração de dados, é o sistema que retorna as regras, não dependendo da racionalidade, muitas vezes limitada do usuário. Assim, a pessoa faz uso dessas regras da forma que for mais atrativa para o seu negócio ou situação.

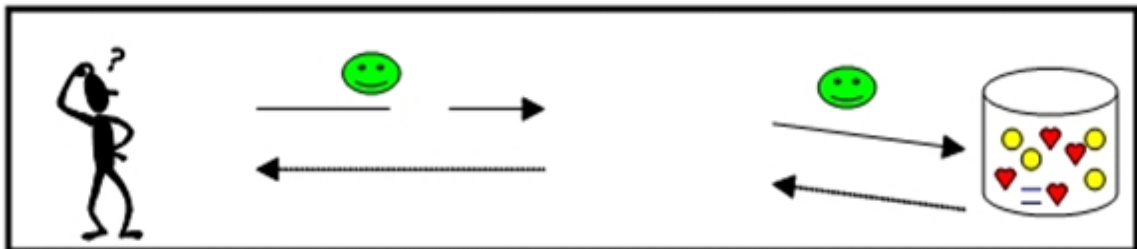


Figura 7 - Busca por informações em sistemas convencionais (SIG, OLAP, Query, etc.)
Fonte: Gonçalves (1999).

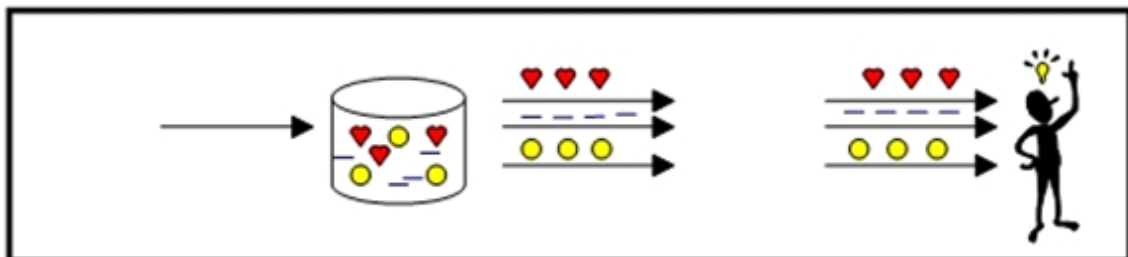


Figura 8 - Busca por informações em sistemas de mineração de dados
Fonte: Gonçalves (1999).

Neste tópico é descrito de forma sucinta algumas tarefas de mineração e com maior ênfase a tarefa de associação. São **tarefas** da mineração de dados

assim: a classificação; a estimativa; a segmentação; a regressão; a agregação (*clustering*); a sumarização; o modelo de dependência; a associação; a análise de sequência, a previsão e a predição; agrupamento por afinidade; redes neurais, dentre outras.

A classificação é tida como uma das tarefas mais comuns da mineração de dados. De acordo com Gonçalves (1999, p.17), a classificação “consiste em examinar os aspectos de um objeto e ligá-los a uma das classes pré-definidas”. Para Lamboia e Pereira (2005, p.5) “várias são as aplicações para este tipo de abordagem, em análise de risco, por exemplo, o objetivo pode ser o de classificar um potencial cliente entre as classes de excelente, bom ou mau pagador”. Gonçalves (1999, p.17) classifica em: atribuir palavras-chave a artigos jornalísticos; classificar pedidos de créditos como de baixo, médio e alto risco; determinar que número telefônico corresponde ao fax; esclarecer pedidos de seguro fraudulentos.

De acordo com Lamboia e Pereira (2005), classificação é a “estratégia que consiste na busca por uma função que consiga mapear (classificar) uma determinada ocorrência em uma dentre um conjunto finito e pré-definido de classes”, ou seja, analisa um conjunto de dados e constrói um modelo para cada classe baseado na característica de dados; na análise de riscos é usada por exemplo, com o objetivo de classificar um potencial cliente entre as classes de mau, bom ou excelente pagador.

Para Baraniuk (2009, p. 149), tal método ajuda na previsão de determinado resultado, ou seja, no método de classificação são procuradas semelhanças entre os dados que estão disponíveis sobre cada indivíduo, formando assim, grupos com comportamentos e características que lhe são peculiares. Podem ser as árvores de decisão indutivas, os modelos Bayesianos, os modelos baseados em regras (*rule-based*), a classificação por *backpropagation* e o modelo de *Support Vector Machines*, entre outros. Uma das técnicas de classificação mais difundidas é a de regras de decisão ou método de indução da estrutura de decisão. Ademais, Baraniuk (2009) cita como ferramenta amplamente utilizada o modelo C5.0, que inclusive foi utilizado por ele em sua pesquisa. O modelo C5.0 é uma ferramenta que permite o trabalho com entradas de dados numéricas ou nominais, desde que o objetivo do modelo seja uma variável nominal.

Comenta Baraniuk (2009) que os modelos de classificação são amplamente utilizados em mineração de dados. Dado tal aplicação, tais modelos podem ser

comparados e avaliados de acordo com alguns critérios, quais sejam: acurácia; velocidade; robustez; escalabilidade; e interpretabilidade. A acurácia de um classificador refere-se à habilidade de um modelo em classificar corretamente e previamente um dado ainda não observado (HAN e KAMBER, 2006, p. 290).

A acurácia também pode ser referida como a taxa de reconhecimento do classificador. A velocidade refere-se aos custos computacionais envolvidos para gerar e utilizar o classificador. Já a robustez se define na habilidade do classificador em realizar a predição correta considerando dados com ruído ou dados faltantes. A escalabilidade está relacionada à habilidade de construir o classificador eficientemente, utilizando-se de grande quantidade de dados. Por fim, a interpretabilidade refere-se ao nível de entendimento e de novas revelações que o classificador oferece. A interpretação, por ser subjetiva, é de difícil avaliação (BARANIUK, 2009).

Associação, por sua vez, salienta Lamboia e Pereira (2005), “consiste em identificar fatos que possam ser direta ou indiretamente associados. Esta estratégia é geralmente usada em aplicações onde se busca identificar itens que possam ser colocados juntos em um mesmo pacote de negociação”, ou seja, desenvolver estratégias associando certos itens a determinados tipos de clientes. Exemplo:

[...] a constatação de que vendas de cerveja e de fraldas descartáveis aumentam às sextas-feiras pode levar à construção de uma hipótese que associe ambos os itens a um tipo especial de cliente. Associação também pode ser usada para avaliar a existência de algum tipo de relação temporal entre os itens constantes de uma base de dados (LAMBOIA e PEREIRA, 2005, p.20).

Os métodos de associação “são ferramentas em mineração para identificar padrões de combinações nos dados, permitindo, como exemplo, verificar a possibilidade de uma pessoa adquirir um determinado produto dado o seu histórico de aquisições” (Baraniuk 2009). Baraniuk (2009), abre parênteses sobre o aplicativo usado por ele na pesquisa:

O aplicativo **Clementine** (SPSS, 2008), a ser utilizado nesta pesquisa, disponibiliza os modelos *Generalized Rule Induction*, Apriori (AGRAWAL e SRIKANT, 1994) e CARMA (HIDBER, 1999).

O modelo de **Indução Geral** (*Generalized Rule Induction* - GRI) busca descobrir regras de associação nos dados. O GRI extrai as regras a partir de um índice que contém as probabilidades e a precisão das regras apresentadas. Este

modelo considera entradas de dados numéricas e por categorias, mas a resposta é apresentada na forma categórica.

O modelo **Apriori** (AGRAWAL e SRIKANT, 1994) extrai um conjunto de regras dos dados, identificando as regras com maior conteúdo de informações. Este modelo requer que os dados de entrada e saída sejam categóricos.

O modelo **CARMA** (HIDBER, 1999), abreviatura de *Continuous Association Rule Mining Algorithm*, extrai um conjunto de regras dos dados sem a necessidade de que sejam especificados os dados preditores ou de previsão. Uma diferença entre os demais métodos, o modelo CARMA permite visualizar regras de associação tanto para os itens antecedentes, quanto os conseqüentes.

Agregação, por sua vez, comenta Lamboia e Pereira (2005, p.6), consiste “na busca de similaridades entre os dados tal que permita definir um conjunto finito de classes ou categorias que os contenha e os descreva”. “Esta tarefa agrupa informações homogêneas de grupos heterogêneos entre os demais e aponta o item que melhor representa cada grupo, também conhecido como *clusters*” (Lamboia e Pereira 2005). Como exemplo em determinadas cidades, o banco verificou a concentração de clientes por região e/ou bairro. Além disso, verificou a natureza das operações que esses clientes realizavam, se eram usuários mais freqüentes de agências, caixas eletrônicos ou Internet. Dessa forma, pôde distribuir novos caixas eletrônicos em regiões onde a oferta desse serviço era pequena, mas havia uma grande quantidade de clientes com potencial para o uso do serviço (CÔRTES, 2004, p.137).

Já, a predição acaba envolvendo uma componente temporal, ou seja, representa aquela classe de problemas nos quais se tem interesse em prever o comportamento ou valor futuro de uma determinada variável com base em valores anteriores desta mesma variável (mono-variável) ou em valores anteriores da variável de interesse e de outras variáveis (multi-variável), sugere Lamboia e Pereira (2005, p.8)

Isto quer dizer que, nessa tarefa há a análise de um grande conjunto de dados de séries temporais com o fim de encontrar certas regularidades e características interessantes.

Afirma Gonçalves (1999, p.18), que a previsão se assemelha com classificação ou estimativa, “exceto pelo fato de que os registros são classificados

de acordo com algum comportamento futuro previsto ou valor futuro estimado”. Alguns exemplos de previsão citados pelo autor são: previsão da quantia de dinheiro que um cliente utilizará caso seja oferecido a ele um certo limite de crédito; previsão de quais clientes sairão nos próximos seis meses; previsão de quais assinantes de telefone usariam um serviço extra, como segmentação por telefone, CPA, ou redirecionamento de ligação.

A estimativa, por sua vez, a partir de algum dado de entrada, é usada para “estipular um valor a uma variável contínua desconhecida, tal como renda, altura ou limite de cartão de crédito” diz Gonçalves (1999).

A segmentação também tida como uma das tarefas do *Data Mining*, busca “realizar a separação em uma grande população de sub-populações com comportamentos semelhantes. Dentro dessas sub-populações, vários tipos de previsões são possíveis” (LAMBOIA e PEREIRA, 2005). Um exemplo de sua aplicação é a segmentação de clientes a partir da descoberta de grupos de clientes que apresentam características de comportamento semelhantes. As demais tarefas são também comentadas por Lamboia e Pereira (2005, p. 5), sendo:

- **Regressão:** Consiste na busca por uma função que represente, de forma aproximada, o comportamento apresentado pelo fenômeno em estudo. A forma mais conhecida de regressão é a linear, por exemplo, uma reta que minimiza o erro médio entre todos os valores considerados, mas também pode ser não linear.
- **Sumarização:** Envolve métodos que encontram uma descrição compacta para um subconjunto de dados. As técnicas de sumarização são freqüentemente aplicadas à análise exploratória interativa de dados e geração automatizada de relatórios.

Aplicações podem ser de tabular o resultado e o desvio padrão **para todos** os atributos de conjunto de dados, descobrir dependência funcional entre as variáveis, realizar a derivação de um sumário de regras, etc.

Lamboia e Pereira (2005, p.7) explicam ainda tarefas como o modelo de dependência e análise de sequência: O Modelo de **Dependência** descreve as dependências significativas entre as variáveis. Os modelos de dependência existem

em dois níveis: estruturais, cujas variáveis são localmente dependentes entre si; quantitativos os quais especificam a solidez da dependência usando uma escala numérica. Geralmente um modelo utiliza os dois tipos de dependências, um para encontrar a relação e outra para quantificá-la. Estes modelos são utilizados em áreas como desenvolvimento de probabilidades em sistemas especialistas da área médica, recuperação de informação e modelagem de genes humanos.

Na **análise de seqüências** uma variante de problema de regras de associação é a análise de seqüencial, onde as regras encontradas entre as relações podem ser usadas para identificar seqüências interessantes, que serão utilizadas para prever acontecimentos subsequentes.

As técnicas de mineração são importantes, pois conseguem desvelar padrões anteriormente ocultos no banco de dados, trazendo à luz informações valiosíssimas para o usuário. Gonçalves (1999, p. 5) comenta que ao encontrar padrões e relacionamentos ocultos nestes bancos de dados, trazem à tona conhecimentos não apenas sobre os objetos registrados, mas também, se a base de dados é um espelho fiel, sobre o mundo real registrado no banco de dados.

Agrupamento por afinidade ou análise da cesta de supermercado é a tarefa de agrupar itens por afinidade, ou seja, quais são as “coisas que vão juntas numa cesta ou carrinho de supermercado, por exemplo” (GONÇALVES, 1999, p.5). Ademais, o agrupamento por afinidade também é usado na identificação de vendas de pacotes de produtos ou também de serviços e produtos. Exemplo: “As cadeias de varejo usam esta técnica para planejar a disposição dos produtos nas prateleiras das lojas ou em um catálogo, de modo que os itens geralmente adquiridos na mesma compra sejam vistos próximos entre si” (GONÇALVES, 1999).

Já segmentação tarefa de segmentar uma população heterogênea em um número maior de subgrupos homogêneos ou *clusters*. No clustering não há classes predefinidas. Na segmentação os registros são agrupados de acordo com a semelhança (Berry 1997).

A segmentação é comumente um dos primeiros passos na análise da mineração de dados. Já sobre a descrição, Gonçalves (1999, p. 19) menciona que algumas vezes o objetivo da mineração de dados é tão somente descrever o que está acontecendo em uma base de dados complicada no intuito de aumentar nosso entendimento sobre as pessoas, produtos ou processos que produziram os dados.

Existem outras técnicas mais específicas da mineração de dados: análise de seleção estatística, MBR³, algoritmos genéticos, detecção de agrupamentos, análise de vínculos, árvores de decisão e indução de regras e redes neurais artificiais.

A análise de seleção estatística é “uma forma de agrupamento usada para encontrar grupos de itens que tendem a ocorrer em conjunto em uma transação ou seleção estatística”. Sobre MBR, algoritmos genéticos, detecção de agrupamentos, análise de vínculos e árvores de decisão e indução de regras, salienta Gonçalves (1999, p. 21):

O MBR (*Memory Based Reasoning*) ou raciocínio baseado em memória é uma técnica que usa exemplos conhecidos como modelo para fazer previsões sobre exemplos desconhecidos (Harrison, 1998). [...]
Os algoritmos genéticos aplicam mecanismos de seleção genéticos e naturais para uma busca usada para encontrar conjuntos de parâmetros ótimos que descreve uma função preditiva.
[...]
Detecção de agrupamentos [...] Harrison (1998) define esta técnica como a construção de modelos que encontram registros de dados semelhantes.

Segundo Harrison (1998) essas reuniões por semelhança são chamadas grupos (*clusters*). Na percepção de Berry e Linoff (1997) a detecção de agrupamentos trata-se de mineração de dados indireto, uma vez que a meta é encontrar similaridades não conhecidas previamente.

Harrison (1998) também discorre sobre a análise de vínculos salientando que esta segue as relações entre registros para desenvolver modelos baseados em padrões nas relações. O mesmo autor salienta que como técnica de mineração de dados, a análise de vínculos não é muito compatível com a tecnologia de bancos de dados relacionais. A maior área onde é aplicada, segundo ele, é a área policial, onde pistas são ligadas entre si para solucionar crimes.

Conforme Gonçalves (1999, p.23), as árvores de decisão são usadas para a mineração de dados direta, mais precisamente para a classificação. Já para Berry e Linoff (1997), as árvores de decisão são um modelo poderoso produzido por uma classe de técnicas que inclui árvores de regressão e de classificação e indução qui-quadrado automática. Harrison (1998) identificou como uma das principais vantagens das árvores de decisão a facilidade de explicação de seu modelo, devido a sua forma de regras explícitas.

³ MBR (*Memory Based Reasoning*) ou raciocínio baseado em memória.

Gonçalves (1999, p. 23) comenta que:

O conceito-chave das **redes neurais** é a utilização de dados na criação de bases de conhecimentos. As redes neurais, ao contrário dos sistemas especialistas não precisam de um especialista para a criação da sua base de conhecimentos. Não trabalha com regras, sua aquisição é feita automaticamente a partir de exemplos coletados em bancos de dados.

Quando a questão é qual técnica usar, não há uma que possa resolver todos os problemas da mineração de dados. É importante que se conheça as técnicas para que a melhor abordagem possa ser escolhida de acordo com os problemas apresentados. Para que o processo decisório seja coerente e atenda os objetivos, bem como, para que o tomador de decisão se utilize das ferramentas disponíveis, é preciso conhecimento não só prático, mas teórico. Assim, é possível a identificação de qual técnica de TI utilizar, de acordo com a ocasião.

Baraniuk (2009, p. 1) defende o uso do *Data Mining* na disponibilidade das informações de cada cliente, pois isto traz amplas possibilidades de atuação para a empresa em atuar em nichos de mercado cada vez menores, onde o nível individual seria o limite. Uma das métricas utilizadas para o monitoramento dos clientes em nível individual é o Valor do Cliente no Tempo, que representa o valor presente da previsão do fluxo de caixa a ser gerado por um cliente durante o período futuro de relacionamento com a empresa. Este parâmetro pode auxiliar os gestores na escolha de ações de relacionamento com os clientes, tanto em nível individual como agregado.

Por assim dizer, o autor antenado com as mudanças dos processos e serviços, apresenta em sua tese a preocupação das empresas, sobretudo, a partir da redução dos custos em informática, com o armazenamento do conjunto de informações dos clientes em sua base de dados, refletindo em um aumento de produtos – leia-se produtos personalizados em muitos dos casos e serviços, e o desenvolvimento de novas ferramentas para gerenciar esses produtos e clientes.

No que diz respeito aos serviços, pode-se comentar que hoje uma das grandes armadilhas é lançar serviços sem estudos prévios adequados, levados pela pressão dos concorrentes ou pela dificuldade de mensurar a resposta dos consumidores aos novos serviços. O setor de TIC (Tecnologia de Informação e Comunicação) apresenta muitos casos em que isso acontece, ou seja, as empresas

têm custos baixos para o lançamento de novos produtos, mas tampouco conhece o perfil do consumidor.

Assim, conforme pontua Baraniuk (2009), a grande quantidade informacional sobre os usuários, armazenadas nos bancos de dados, fornece novos desafios à Administração, mas também novas oportunidades, e para isso são necessárias novas abordagens e modelos levando em consideração esta nova perspectiva.

Situação semelhante acontece com o setor de telecomunicações. Vários dos serviços disponibilizados são de alta complexidade, dessa forma, há dificuldade de compreensão das novas formas de aplicação, bem como não é possível prever o comportamento dos usuários, nem conhecer os riscos associados, já que não há parâmetros para comparação. Nessa perspectiva, surge a procura por métricas, a partir da qual se pode avaliar os consumidores e desenvolver novas ferramentas de apoio decisório diante do volume de informações disponíveis.

Baraniuk (2009) apresenta o **valor do cliente no tempo** (CLV), que é um dos métodos de avaliação do consumidor a nível individual. A esse respeito diz o autor que o CLV representa o valor presente do fluxo de caixa a ser gerado por cada cliente no período de relacionamento. Levando-se em conta a previsão de receita a ser gerada por cada cliente e, também, os gastos relacionados à prestação de serviços e à retenção. Desta forma, as empresas têm a possibilidade de gerenciar ações junto aos clientes, de modo a otimizar a rentabilidade tanto a curto como a longo-prazo.

Assim, pode-se frisar que o CLV atualmente é um método bastante relevante no que diz respeito à segmentação, seleção e retenção de clientes. Além do mais, o modelo auxilia na mensuração em termos monetários dos lucros auferidos a partir da angariação e fidelização dos clientes, ou seja, pode calcular o valor econômico do cliente.

Para Seixas (2007, p. 4), o valor de um cliente e, conseqüentemente, da empresa poderão aumentar se houver uma redução do custo inicial de aquisição do cliente ou uma maior contribuição ao longo da vida do cliente. Em outras palavras, o cliente é um ativo, de certa forma, mais valioso que os ativos tangíveis do balanço da empresa. Os ativos da folha de balanço podem ser comprados, mas os clientes não se “compram” com tanta facilidade. Um dos segredos do *Customer Lifetime Value* é a informação disponibilizada para a tomada de decisão sobre o melhor relacionamento com cada cliente. Lembrando que diferentes ações resultam em

diferentes resultados e que nestes casos o modelo *Customer Lifetime Value* pode ser muito útil ao permitir a escolha das melhores ações no sentido de aumentar o tempo-vida dos clientes. Por este fato o *Customer Lifetime Value* tem uma “abordagem a longo prazo”.

O modelo CLV é composto por alguns elementos:

- **Fontes Externas** – que dizem respeito às bases de dados externas à empresa estudada, conseguida através de censos ou afins, mas que permitem caracterizar uma população. “Outro tipo de fonte externa são as pesquisas de campo, ou *surveys*, que são aplicadas sobre uma amostra de clientes, com o intuito de obter maiores informações sobre o comportamento deles” (Baraniuk 2009).
- **Fontes Internas** - compostas pela base de dados transacionais⁴; no caso dos serviços de telecomunicações são as informações financeiras e de uso dos serviços, registradas automaticamente. As informações obtidas pela equipe de suporte de serviços a partir do contato com usuários seja por telefone, *e-mail* ou visita direta - formam a base de dados de relacionamentos. (Baraniuk 2009).
- **As atividades de Mineração de Dados** - são compostas por técnicas para preparar os dados disponíveis e gerar o indicador do CLV. Já o Valor do Cliente no Tempo, constitui-se como a principal variável a ser obtido no modelo, identificando-se os clientes por faixa CLV. Para a interpretação dos dados, o pesquisador optou por utilizar o modelo de regras de decisão gerados pelo algoritmo C5.0. (Baraniuk 2009).
- **A Coleta de Informações** - de aquisição, retenção e expansão de uso por sua vez - “tem por objetivo realimentar as bases de dados, de

⁴ As bases de dados transacionais contém todas as informações operacionais da empresa, tais como: cadastros de clientes, produtos, fornecedores, pessoal, representantes, movimentos de faturas, notas fiscais, documentos de cobrança, entradas e saídas de estoque, lançamentos financeiros, registros fiscais, lançamentos contábeis, eventos de pagamento, registro de serviços, atendimento a clientes, etc. Resumindo, são todas as informações históricas armazenadas ao longo do tempo de vida da empresa, geralmente imensos volumes de dados e muitas espalhadas em várias fontes de dados. Disponível em: < http://utilsoft.com.br/utilbi/bi_caract_tecnicas.htm#Base_de_Dados_transacional>. Acesso: 23 mar. 2010.

modo a permitir o aprimoramento do modelo, possibilitando a evolução, o ajuste e a avaliação das ações executadas”. (Baraniuk 2009).

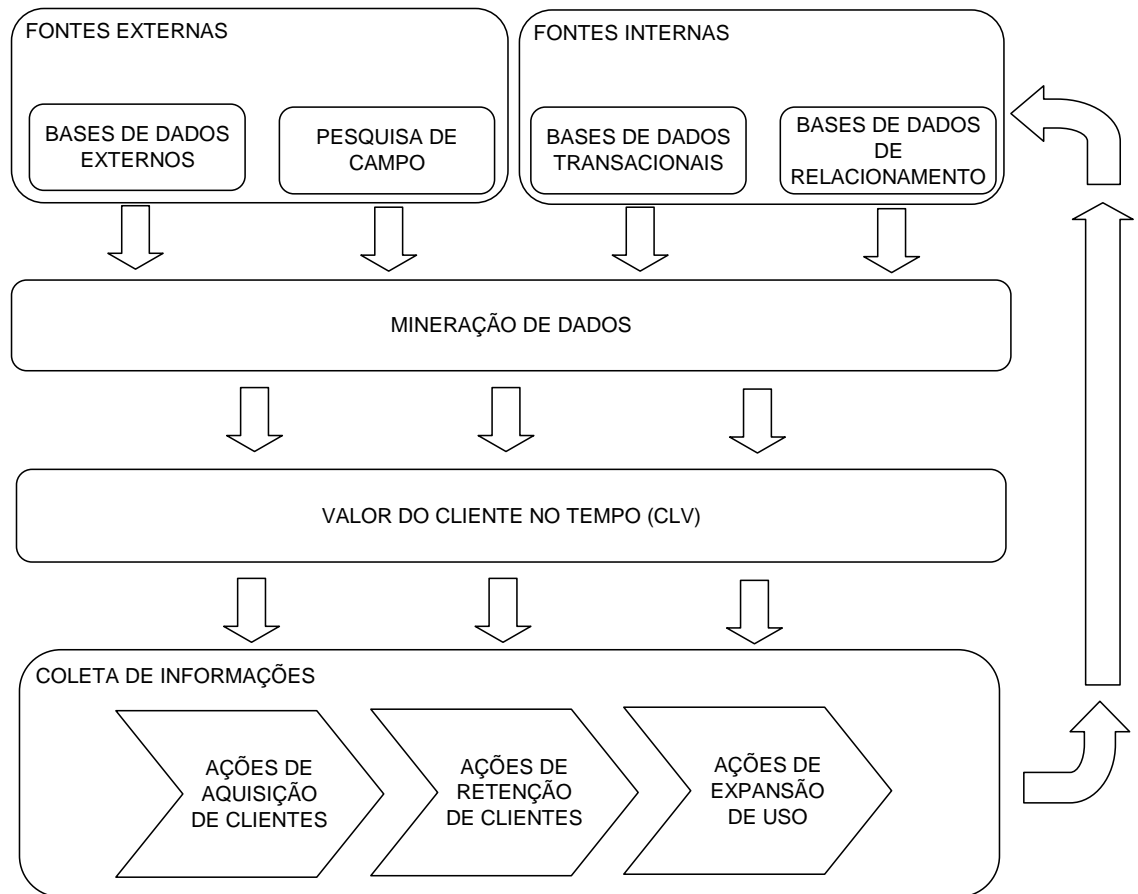


Figura 9: Modelo Integrado de Gerenciamento pelo Valor do Cliente no Tempo
Fonte: Baraniuk (2009).

Os modelos usados por Baraniuk (2009) foram os da regra de decisão e de árvore de decisão, que segundo ele, são mais fáceis de compreensão e interpretação. Considera o autor para execução o uso de modelos simbólicos por permitirem uma melhor compreensão das relações entre as variáveis, sendo utilizado o algoritmo de regras de decisão C5.0.

A pesquisa de Baraniuk (2009) foi dividida em duas fases, sendo uma exploratória, na qual se buscou maiores informações sobre a empresa pesquisada, sobre os consumidores e serviço, objeto de estudo, a partir de procedimentos qualitativos; e a outra fase de modelagem com procedimentos quantitativos, a partir das técnicas de mineração de dados como referência a metodologia CRISP-DM para estruturação das atividades. A pesquisa seguiu as seguintes etapas:

- Estudo dos dados disponíveis
- Estudo das formas de enriquecimento dos dados
- Aplicação de técnicas que permitem encontrar padrões que expliquem as variáveis-objetivo, utilizando o algoritmo de modelagem C5.0, disponível no aplicativo *Clementine* (SPSS, 2008), para a construção de um modelo simbólico de raciocínio
- Seleção dos padrões que melhor se ajustam para o modelo
- Validação do modelo obtido.

Há um capítulo dedicado integralmente à mineração de dados, no qual o pesquisador disserta sobre a Metodologia CRISP-DM. Tal modelo (*Cross Industry Standard Process for Data Mining*), segundo o autor, foi proposto pelo grupo de mesmo nome, fundado em 1996 pela *Daimler Chrysler*, SPSS e NCR. A figura 10 apresenta o modelo hierárquico que diz respeito a esta metodologia. São quatro seus níveis de abstração: a) fase – no primeiro nível, o processo de mineração de dados é organizado em fases, na qual serão realizadas atividades genéricas em um segundo nível; b) atividade genérica – devem ser as mais completas e genéricas possíveis, procurando cobrir todas as situações possíveis da mineração de dados; c) atividade específica – devem ser conduzidas em situações específicas; e d) instância de processo – na qual são registradas ações, as decisões e os resultados da mineração de dados.

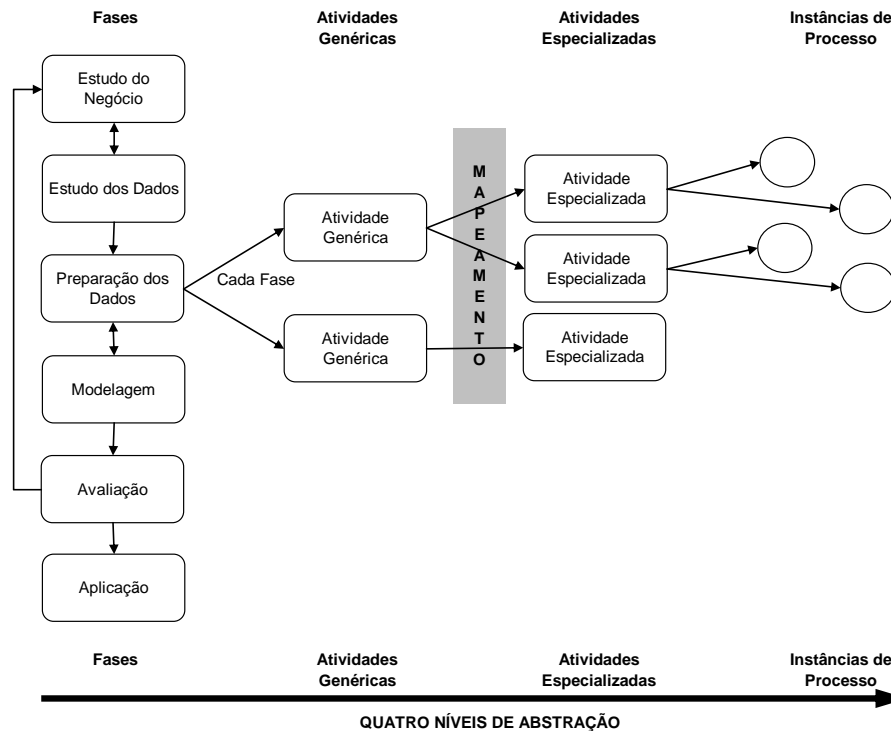


Figura 10: Metodologia CRISP-DM
 Fonte: Sumathi e Sivanandam (2006, p. 702).

O quadro 3 apresenta as fases e as atividades genéricas do modelo CRISP-DM.

Entendimento do Negócio	Entendimento dos Dados	Preparação dos Dados	Modelagem	Avaliação	Aplicação
Verificar os Objetivos do Negócio	Coleta Inicial de Dados	Preparação dos Dados	Seleção da Técnica de Modelagem	Avaliação dos Resultados	Planejamento dos Relatórios
Acesso aos Dados	Descrição dos Dados	Seleção dos Dados	Geração de Modelo de Teste	Revisão do Processo	Planejamento do monitoramento e manutenção do modelo
Definição do Objetivo da Mineração de Dados	Exploração dos Dados	Limpeza dos Dados	Construção do Modelo	Definição das próximas ações	Produção de Relatório Final
Planejamento do Projeto	Verificação da Qualidade dos Dados	Construção dos Dados	Implementação do Modelo		Documentação do Projeto
		Integração dos Dados			
		Formação dos Dados			

Quadro 3: Fases e atividades genéricas do modelo CRISP-DM
 Fonte: BARANIUK (2009) adaptado de Sumathi e Sivanandam (2006, p. 766).

Baraniuk (2009, p. 255) explica sinteticamente cada um dos estágios, sendo:

A metodologia CRISP-DM segue um modelo de seis estágios, [...] sendo definidos por: entendimento do negócio; entendimento dos dados; preparação dos dados; modelagem; avaliação; e aplicação. [...] O estudo do negócio consiste na busca de informações, como: dados estatísticos, estudos realizados e modelos já construídos sobre o tema a ser abordado. O entendimento dos dados consiste na interpretação de cada variável disponibilizada, bem como a preparação destas variáveis para serem processadas nos programas de mineração de dados. A preparação dos dados consiste em operações de manipulação matemática dos dados, incluindo seleção, amostragem, combinação, seleção e preparação de dados agregados. A etapa de modelagem consiste em ferramentas de análise, buscando o entendimento do problema que se pretende resolver. A avaliação envolve a interpretação e a visualização dos resultados obtidos pelos processos de mineração de dados, utilizando-se de tabelas, gráficos, histogramas e matrizes, entre outras ferramentas. A apresentação envolve a utilização das informações na empresa, que pode ser na implementação de um modelo a ser utilizado pela organização.

Pretende-se, aqui, fazer um detalhamento sobre o processo de modelagem, segundo Baraniuk (2009) que seria uma das partes mais interessantes para a **pesquisa proposta por este trabalho**, para depois dar prosseguimento à análise das próximas fases.

O modelo, de acordo com Baraniuk (2009, p.147), “procura captar os elementos mais importantes de uma realidade complexa, possuindo limitações, uma vez que são utilizados elementos percebidos pelo construtor do modelo e pelas pessoas envolvidas no sistema real em estudo”. No que diz respeito à sua utilização, é possível relacionar três intenções de uso para o modelo, quais sejam: descritivos – orientados para a descrição do processo decisório ou para compreensão de outros fenômenos como satisfação e lealdade da marca; preditivos - objetivam fazer a previsão dos eventos futuros, indicando tendências ou respondendo questões relacionados ao planejamento estratégico, como lançamento de um novo produto ou serviço por exemplo⁵; e o modelo normativo ou prescritivo – que definem ações a partir de situações que já ocorreram.

Baraniuk (2009, p.147) resume que os modelos descritivos auxiliam o entendimento do “por quê” dos fenômenos em estudo; os modelos preditivos respondem a questões sobre o futuro, se determinada ação for tomada; e, o modelo

⁵ para Baraniuk tal modelo, é muitas vezes insatisfatório por se basear o comportamento futuro em padrões já existentes

normativo, indica as ações que devem ser tomadas mediante as situações encontradas.

A composição da fase de modelagem é a seguinte: seleção de modelo, definição do modelo de teste, construção do modelo e avaliação do modelo.

Assim, no entendimento de Baraniuk (2009, p.147), a primeira ação, seleção e modelo, incide em escolher as técnicas a serem utilizadas, identificando os requisitos de dados e questões quanto ao resultado a ser obtido para cada modelo. A segunda ação, geração de modelo de teste, consiste em definir o procedimento a ser utilizado para validar e verificar a qualidade do modelo. A ação seguinte, de construção do modelo, consiste na implementação do modelo e respectiva simulação. Ao final é realizada a ação de avaliação do modelo, que consiste na definição do procedimento de teste e do critério de sucesso a ser utilizado.

Na fase exploratória, Baraniuk (2009) procedeu com a coleta de dados secundários que envolveu a busca de informações relacionadas ao serviço VoIP; já os dados primários foram obtidos por meio de reuniões e visitas junto à empresa prestadora de serviço VoIP, bem como, com gestores e um grupo de consumidores, clientes da empresa em estudo, utilizando-se de entrevistas semi-estruturados.

Já para a fase quantitativa, os dados secundários foram obtidos a partir de dados econômicos e demográficos do IBGE (Instituto Brasileiro de Geografia e Estatística); MTE (Ministério do Trabalho) e IPARDES (Instituto Paranaense de Desenvolvimento Econômico e Social). Frisa ainda o pesquisador:

Especificamente, verificou-se a quantidade de empresas e trabalhadores para cada município brasileiro, utilizando-se de informações disponíveis na base de dados da Relação Anual de Informações Sociais (RAIS), disponibilizadas pelo Ministério do Trabalho (MTE, 2009), em que os dados são disponibilizados gratuitamente na *home Page* para uso de fins acadêmicos.

Como dados primários, foram utilizadas as informações disponibilizadas pela empresa, incluindo as informações contidas no *Google Analytics* (GOOGLE, 2008) do serviço em estudo e, também, amostras de dados transacionais disponíveis na base de dados da empresa. [...] O *Google Analytics* é uma ferramenta de monitoramento e de análise de acesso a página da *Internet*. As páginas que são monitoradas pelo *Google Analytics* registram o número de visitas realizadas, apresentando os dados agregados por município. Também são disponibilizadas informações relacionadas ao conjunto de páginas selecionadas, horário de acesso e tempo de acesso. As informações podem ser combinadas em diversas opções de gráficos e figuras, gerando relatórios por dia, semana ou mês. Cabe ressaltar que nesta ferramenta não são disponibilizadas as informações individuais dos usuários, sendo o município o menor nível de agregação obtido (BARANIUK, 2009, p.147).

Os clientes atendidos entre abril de 2006 e fevereiro de 2009 são contemplados pelo *Google Analytics*, sendo possível visualizar a área geográfica em que os usuários acessaram a página de serviço da *Internet*.

Descrição	Google Analytics (Google, 2008)	Base de Dados da Empresa
Foco das informações	<ul style="list-style-type: none"> – Busca de informações sobre o serviço – Recarga do serviço 	<ul style="list-style-type: none"> – Uso do serviço – Receita do Serviço
Tipo de Informação	<ul style="list-style-type: none"> – Número de visitas à Página na <i>Internet</i> – Número de Recargas – Tempo de Acesso 	<ul style="list-style-type: none"> – Receita Mensal – Recarga Mensal – Plano de Serviço – Consumo de ligações de entrada e saída
Origem dos Dados	<ul style="list-style-type: none"> – Monitoramento do acesso pelo cliente à Página na <i>Internet</i> 	<ul style="list-style-type: none"> – Uso dos serviços pelo cliente, registrado pelo equipamento que executa a chamada telefônica
Nível de Agregação (Glanularidade)	<ul style="list-style-type: none"> – Por Município 	<ul style="list-style-type: none"> – Individual (Conjunto 1) – Por Município (Conjunto 2)
Ferramentas de Análise Apropriadas	<ul style="list-style-type: none"> – <i>Web Mining</i> 	<ul style="list-style-type: none"> – <i>Data Mining</i>
Pontos Fortes	<ul style="list-style-type: none"> – Disponibilizada a informação sobre a totalidade dos usuários – Não há erros de coleta de dados, pois a aquisição é automática – Disponibiliza a região geográfica de acesso 	<ul style="list-style-type: none"> – Permite a elaboração de modelos a nível individual – Informação de município preenchida pelo próprio usuário
Pontos Fracos	<ul style="list-style-type: none"> – Municípios com diferentes perfis podem estar agrupados em uma única localidade 	<ul style="list-style-type: none"> – Estudo realizado por amostras – Necessidade de filtragem para a identificação da localidade

Quadro 4: Fontes de Informações Disponibilizadas pela Empresa em Estudo
Fonte: Baraniuk (2009).

Dois conjuntos de informações foram fornecidos a partir da base de dados da empresa, quais sejam, o primeiro “contempla a receita e o número total de clientes em cada localidade, sendo a menor glanularidade o município”. Já o segundo conjunto contempla “dados individuais de uma amostra de 6.108 usuários com o histórico de receita, recarga e uso mensal, abrangendo o período de novembro de 2005 a outubro de 2008, não tendo sido disponibilizada a informação da localidade destes usuários de modo desagregado”. O *Google Analytics* (GOOGLE, 2008) e a base de dados da empresa foram as fontes que disponibilizaram a localidade dos clientes, mês a mês.

Baraniuk (2009) apresenta primeiramente o resultado das reuniões e entrevistas. Diz ele que foram realizadas “sete reuniões em conjunto com a equipe prestadora de serviços, no período de outubro de 2008 a fevereiro de 2009”. Ademais, objetivando melhorar o “entendimento do uso serviço VoIP pela perspectiva do consumidor, foram realizadas três entrevistas com grandes usuários da cidade de Curitiba”. Os resultados obtidos foram os seguintes:

Item	Usuário A	Usuário B	Usuário C
Local	Curitiba	Curitiba	Curitiba
Tempo de Entrevista	90 min	27 min	32 min
Área de Atuação	– Atacadista de Equipamentos de Automação Comercial	– Cobrança	– Fabricante de Equipamentos de Alarmes
Modo de Entrada do Serviço	– O responsável de TI da empresa fez o levantamento	– A operadora ofereceu o serviço	– O consultor de TI da empresa ofereceu o serviço
Item	Usuário A	Usuário B	Usuário C
Responsável pela Decisão	– Responsável pela TI	– Responsável pela área administrativa e financeira	– Responsável pela área de TI e telefonia.
Conhecimento em Tecnologia de Telefonia	– Avançado	– Básico	– Intermediário
Uso do VOIP	– Tele atendimento	– Cobrança	– Tele atendimento
Tipos de Ligações	– Interurbano Nacional – Entrada e saída de ligações – De <i>Softphone</i> para telefone fixo e celular	– Interurbano Nacional – Saída de ligações – De ATA para telefone fixo	– Interurbano Nacional – Entrada e Saída de Ligações – De <i>Softphone</i> para telefone fixo e celular interurbano
Modo de Uso	– <i>Softphone</i> (principal) – <i>Notebook</i> – ATA – Telefone SIP – Celular (pouco uso)	– ATA	– <i>Softphone</i> (principal) – <i>Notebook</i> – ATA (pouco uso) – Telefone SIP (pouco uso)
Recursos Utilizados	– Central de Atendimento com Computador e <i>Softphone</i> – Acesso Remoto por <i>Notebook</i> – Telefones em outras cidades – <i>Home-office</i> com ATA – Secretária por e-mail – Extrato	– Telefone Convencional com ATA – Extrato	– Central de Atendimento com Computador e <i>Softphone</i> – Acesso Remoto por <i>Notebook</i> – Telefones em outras cidades
Vantagens	– Preço	– Preço	– Preço

Apontadas	<ul style="list-style-type: none"> – Estabilidade – Permitiu a criação de números em outras cidades – Permite a mobilidade – Novos serviços telefônicos 		<ul style="list-style-type: none"> – Estabilidade – Permitiu a criação de números em outras cidades – Permite a mobilidade – Viabilizou atividades de vendas via <i>call-center</i>
Desvantagens Apontadas	<ul style="list-style-type: none"> – Não passa FAX – Forma de pagamento como pré-pago 	<ul style="list-style-type: none"> – Não passa FAX – Há números em que as ligações não são completadas 	<ul style="list-style-type: none"> – Não recebe ligação a cobrar – Há regiões em que a ligação não é completada
Sugestões para Ampliação do Uso	<ul style="list-style-type: none"> – Divulgar para outros usuários 	<ul style="list-style-type: none"> – Compatibilidade com FAX – Reduzir o custo do celular 	<ul style="list-style-type: none"> – Receber ligações a cobrar – Aumentar o número de cidades atendidas
Modo de Entrada do Serviço	<ul style="list-style-type: none"> – O responsável de TI da empresa fez o levantamento 	<ul style="list-style-type: none"> – A operadora ofereceu o serviço 	<ul style="list-style-type: none"> – O consultor de TI da empresa ofereceu o serviço
Responsável pela Decisão	<ul style="list-style-type: none"> – Responsável pela TI 	<ul style="list-style-type: none"> – Responsável pela área administrativa e financeira 	<ul style="list-style-type: none"> – Responsável pela área de TI e telefonia
Conhecimento em Tecnologia de Telefonia	<ul style="list-style-type: none"> – Avançado 	<ul style="list-style-type: none"> – Básico 	<ul style="list-style-type: none"> – Intermediário

Quadro 5: Relação de Entrevistas Exploratórias com Empresas Usuárias de VoIP
Fonte: Baraniuk (2009).

Os equipamentos mais utilizados são os *softphones* – “aplicativos que permitem o uso do computador como um telefone convencional” - e o ATA - Adaptador de Telefone Analógico “que é um dispositivo que permite o uso de um aparelho telefônico convencional à rede de comunicação de banda larga”. O telefone IP (ou SIP) e o telefone via dispositivo móvel celular são equipamentos também encontrados, embora menos comuns.

Baraniuk (2009) observa que os equipamentos da prestadora de serviços estudada permitem a identificação do tipo de dispositivo utilizado pelo usuário VoIP. Esta informação tende a auxiliar a identificar os usuários líderes ou os usuários inovadores.

As entrevistas detectaram a importância da marca, da qualidade do serviço e dos serviços auxiliares no que diz respeito à operadora de serviços de telecomunicações. O principal atributo, no entanto, é o preço, ainda que haja a exigência de que o serviço possua qualidade.

Baraniuk (2009) ainda frisa que há uma relação de confiança entre o usuário e o responsável pela área de TIC organizacional. Notou o pesquisador que a tecnologia VoIP necessita de especialistas no setor para ser implementada, e isso

caracteriza-se como barreiras para seu uso e difusão. Frisa ainda o pesquisador que os diversos modos de uso do serviço VoIP têm permitido a expansão de negócios para os clientes. Finaliza Baraniuk (2009) afirmando que as reuniões e entrevistas permitiram melhor compreensão sobre as características do produto e do serviço, bem como do mercado de VoIP em si.

Baraniuk (2009, p.147) destaca as características do serviço VoIP, sendo: a) serviço pré-pago, em que o usuário necessita adquirir créditos para poder utilizar os serviços; b) a realização de chamadas telefônicas entre usuários do mesmo serviço VoIP é gratuita, só ocorrendo a cobrança de tarifas quando a chamada é realizada para um usuário de outra rede de serviços telefônicos; c) os créditos possuem prazo para utilização, expirando mesmo que não tenham sido utilizados; d) são ofertados diversos pacotes de serviços com diferentes combinações de custos de ligações e de oferta de serviços complementares.

Na etapa quantitativa, o autor apresenta cinco modelos relacionados ao gerenciamento do Valor do Cliente no Tempo - CLV. No primeiro modelo que seria o de Identificação de Clientes de CLV elevado, a pesquisa considerou como CLV usuários com expectativas acima de **R\$ 3.000,00**, considerando-se doze meses após o período de análise.

Neste modelo, os clientes são classificados em níveis de acordo com o potencial de geração de receita. Foram considerados 6.108 usuários da empresa pesquisada, para que em seguida, fossem realizadas diversas atividades, de modo a encontrar os usuários de CLV elevado ou CLV baixo. Não serão apresentados os passos para obtenção do CLV, por fugir do foco de pesquisa. Contudo, o autor avalia o modelo como eficaz para a classificação dos usuários.

O segundo modelo é o de obtenção do CLV, no qual os clientes são classificados em níveis, de acordo com o potencial de geração de receita. Baraniuk (2009) expõe também o modelo auxiliar para a aquisição de clientes, ou melhor, este item estava “voltado à identificação dos Municípios com potencial de adesão de um grande número de usuários de “CLV Elevado”. Tal modelo objetiva produzir informações que possam auxiliar na angariação de novos clientes para o serviço VoIP da empresa em questão. Assim diz o pesquisador, que o modelo servirá para possivelmente identificar os municípios e setores econômicos em potencial para gerar as receitas elevadas. O quadro 6 mostra as diferenças básicas entre as informações adquiridas, sendo:

Descrição	Google Analytics	Base de Dados
Foco das informações	Busca de informações sobre o serviço Recarga do serviço	Uso do serviço Receita do Serviço
Tipo de Informação	Número de visitas à Página na <i>Internet</i> ; Número de Recargas Tempo de Acesso	Receita Mensal Recarga Mensal Plano de Serviço Consumo de ligações de entrada e saída
Origem dos Dados	Monitoramento da Página na <i>Internet</i>	Uso dos serviços, registrado pelo equipamento que executa a chamada telefônica
Nível de Agregado	Por Município	Por Município (Conjunto 2) Individual (Conjunto 1)
Ferramentas de Análise Apropriadas	<i>Web Mining</i>	<i>Data Mining</i>
Pontos Fortes	Disponibilizada a informação sobre a totalidade dos usuários Não há erros de coleta de dados, pois a aquisição é automática Disponibiliza a região geográfica de acesso	Permite a elaboração de modelos a nível individual Informação de município preenchida pelo próprio usuário
Pontos Fracos	Municípios com diferentes perfis podem estar agrupados em uma única localidade	Estudo realizado por amostras Necessidade de filtragem da identificação da localidade

Quadro 6: Diferença entre as informações adquiridas sobre os clientes

Fonte: Baraniuk (2009, p.218).

Baraniuk (2009) verificou que, após a filtragem e a conversão do campo de localidades, as informações da base de dados da própria empresa mostraram-se mais adequadas para a realização do modelo de aquisição de clientes. As informações do *Google Analytics* e da base de dados mostraram-se muito semelhantes; no entanto, o *Google Analytics* fornece a informação da operação de atividade de recarga, enquanto que a base de dados da empresa fornece informações sobre a receita por localidade.

Em uma comparação entre a receita gerada por município e o número de estabelecimentos por Divisão Econômica, utilizando-se os dados da RAIS (MTE, 2009), Baraniuk (2009) observou que o modelo gerado produziu acerto superior a 90% para a classificação do nível de receita, permitindo identificar os setores econômicos mais relevantes para o serviço. A comparação entre a base de dados permite identificar os municípios com maior potencial para o desenvolvimento dos serviços em função do número de estabelecimentos por divisão econômica.

Ademais, o modelo conseguiu identificar as localidades com maior potencial para o serviço.

O modelo de identificação de abandono (*churn*) é outro modelo apontado pelo estudo e objetiva identificar usuários com alto risco de abandono da empresa prestadora de serviços. A partir da realização de simulações, a pesquisa apenas encontrou resultados razoáveis de predições até o quarto mês. Após este período, o modelo não demonstrou forte capacidade de avaliação.

Já o modelo de identificação de abandono de clientes como abordagem alternativa, levava em consideração a possibilidade de abandono dos clientes com receita médica superior a **R\$ 500,00**. “O que diferencia este modelo do anterior é o fato deste modelo estar focado em grandes usuários e por buscar e identificar padrões de comportamento dos usuários que abandonaram o serviço, possuindo um maior grau de generalização”. A partir deste modelo, o pesquisador conclui que “grandes variações no consumo e a tendência à redução, são indicadores eficazes para identificar usuários com alto risco de abandono”.

No que diz respeito ao modelo de retenção de clientes obtido, a pesquisa verificou que é possível identificar padrões em clientes com maior risco de *churn* a partir do uso da modelagem com mineração de dados.

Tais modelos apresentados pelo pesquisador ressaltam os benefícios obtidos através do uso das técnicas de mineração de dados, a partir das quais os dados que até então não possuíam nenhuma significância, podem se transformar em informações valiosíssimas, que certamente irão auxiliar o gestor no gerenciamento de clientes e serviços. Diz Baraniuk (2009, p.243):

A utilização de mineração de dados pode auxiliar na elaboração de novos modelos contextualizados a casos brasileiros. Trata-se de uma área de conhecimento ainda incipiente, em que muitos dos estudos são realizados por especialistas em computação. Com a oferta de novas ferramentas de fácil utilização, associada à disponibilidade de grande quantidade de dados dos clientes armazenados em bases de dados eletrônicas, abrem-se novas possibilidades para a realização de modelos a serem utilizados por pesquisadores e gestores de empresas.

Tal observação do pesquisador é de extrema relevância, pois muitos gestores não utilizam a técnica de mineração de dados por desconhecimento desta possibilidade oferecida pela TIC ou até mesmo pelo custo, já que se vê obrigado a investir em profissionais do ramo computacional, dada a dificuldade em manusear a ferramenta. Ademais, muitos ignoram os benefícios da mineração de dados e assim

não conseguem enxergar motivos para investir na técnica, encarando-a como custo e não como investimento.

Min (2006, p.1), é outro pesquisador que disserta sobre a mineração de dados. Min ressalta que para se manter competitivo, os supermercados necessitam desenvolver uma estratégia viável da retenção de cliente. Desde que a chave para o bem sucedido desenvolvimento do negócio esteja apoiada na estratégia de relação entre cliente e gerência, os supermercados devem identificar as maneiras mais rentáveis de construir e manter um relacionamento leal com o cliente. Em um esforço para ajudar os supermercados a compreender o comportamento dos seus clientes/consumidor e as maneiras de reter clientes valiosos, propõe-se a técnica da mineração de dados (prospecção de dados/*Data Mining*). Usando os exemplos de concessões (franquias) de supermercados no sudeste dos Estados Unidos, este artigo ilustra/exemplifica a utilidade da técnica de mineração de dados (prospecção de dados/*Data Mining*) por meio da investigação do comportamento dos clientes na compra de alimentos e desenvolvendo os perfis do consumidor leal.

Min comenta da importância de conhecer as preferências de seus clientes, embora saliente que este ato, por si, não gera vantagem competitiva. O que vai gerar subsídios para superar a concorrência é ter o que ele chama de GRC (Gerenciamento de Relações com o Cliente) ou CRM (*Customer Relationship Management*), bem como o desenvolvimento posterior de uma estratégia de retenção do cliente.

Segundo Min (2006, p.1), o GRC se resume numa prática empresarial que tem o objetivo de melhorar o fornecimento de serviços, construir ligações sociais e assegurar a fidelidade dos clientes, consolidando a longo prazo, uma relação de mútuo benefício com clientes selecionados.

Assim, pode-se dizer que o GRC envolve relacionamento e pretende que ele seja duradouro entre clientes e empresa. Normalmente seu foco, são aqueles clientes que comprem uma grande quantidade de produtos e são fiéis a determinado supermercado por longo período de tempo. Assim, obviamente quanto mais tempo permanecem em um estabelecimento comercial, leia-se supermercado, mais rentáveis se tornam para esta empresa; motivo pelo qual, todos os esforços de retenção da clientela devem ser direcionados a esta fatia de clientes. O artigo focado neste assunto, faz as seguintes perguntas:

1. Como pode a população consumidora ser segmentada entre consumidor fiel e clientes não confiáveis?

2. Que atributos de serviços são mais importantes para cada cliente?
3. Qual segmento de cliente melhor se enquadra na atual capacidade de serviços do supermercado?
4. Quais clientes devem ser alvo por marketing de nicho ou e campanhas promocionais?

Tais questões, poderão ser respondidas utilizando-se o *Data Mining*. Min (2006) recomenda esta técnica para perfilamento dos clientes, dada a sua habilidade em reconhecer e seguir padrões dentro de uma base de dados.

Para responder a pesquisa, o pesquisador se valeu da seguinte metodologia: os respondentes foram no total de 301 consumidores, compradores de dez diferentes supermercados do sudeste americano entre os períodos de 2001 e 2003, que tenham ido a pelo menos um dos dez supermercados deste estudo – o autor faz uma ressalva frisando que quase 73% dos respondentes salienta que freqüentou um dos dez supermercados pelo menos uma vez na semana. A escolha de tais supermercados se deveu ao fato de serem parecidos em suas características no que diz respeito à concessão de franquias, localização e cortesia de serviços, quais sejam disponibilidade de departamentos especiais e caixas rápidos.

Questionários foram respondidos pelos clientes, que colocaram seus dados demográficos, a frequência de visita ao supermercado, quais eram seus padrões de compras; responderam também qual a importância relativa da qualidade total dos serviços do supermercado, bem como o nível de satisfação baseado na sua experiência enquanto consumidor.

Alguns dados estatísticos apontados pelo pesquisador destaca que a maioria (72.7%) dos participantes disse ter visitado um destes supermercados ao menos uma vez por semana. Uma grande maioria (95.7%) disse ter comprado em um dos supermercados mais de uma vez por mês. Aproximadamente três quartos (74.3%) dos participantes comprou em média mais de dez itens por visita ao supermercado. (MIN, 2006).

Assim, pode-se frisar que grande parte dos clientes respondentes da pesquisa são clientes assíduos nos supermercados e conhecedores da qualidade dos serviços e produtos destes estabelecimentos. Min confessa que tais participantes foram selecionados com o intuito de maximizar respostas em seus questionários. Tais clientes foram abordados nas saídas dos supermercados e incentivados, por meio de cupons de desconto ou por abordagem pura e simples, a responder as perguntas.

Com relação ao instrumento de pesquisa, Min (2006) destaca que tradicionalmente os dados coletados dos questionários são analisados utilizando-se técnicas estatísticas que buscam verificar certas premissas através de uma série de testes hipotéticos, muito embora o teste hipotético possa permitir validar algumas premissas intuitivas, as interferências causais feitas pelo teste não podem ser suficientes para prever exatamente os padrões de comportamento dos consumidores em diferentes setores dos supermercados.

E por isso, o autor utilizou-se para sua pesquisa da técnica *Data Mining*. Como método foi utilizado o da árvore de decisão, de acordo com ele, por seu apelo visual, simplicidade de regras e eficiência na classificação dessas regras. De acordo com Menon e Sharda (1999), as árvores da decisão objetivam fazer a classificação de informações dentro de um infinito número de classes, para posteriormente gerar uma hierarquia de afirmações de “SE \Rightarrow ENTÃO”. Assim, com base em uma série de “SE \Rightarrow ENTÃO” pode-se prever comportamentos de compra e consumidor dos clientes do supermercado. Para construir a árvore da decisão seguiu-se três etapas chave segundo as indicações da figura 11 e descrição a seguir: levantamento de dados, formato de dados e regra de indução.

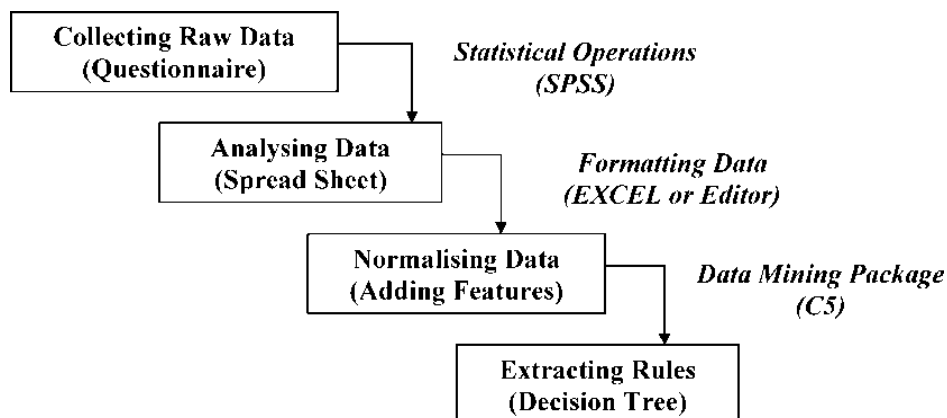


Figura 11: Passos para *Data Mining*
Fonte: Min (2006).

Conforme já mencionado, os dados foram levantados por questionários. Em relação ao formato de dados, para que houvesse exatidão da classificação, “formatou-se séries de dados de maneira que pudessem facilmente ser questionados e então seus padrões poderiam ser especificados em diferentes

categorias”. Com relação à regra de indução, geradas pelas árvores de decisão, pretende-se que sejam facilmente entendidas pela gerência da empresa, podendo ser reaplicados em outros supermercados.

A esse respeito Min (2006, p.6) ressalta que as regras devem dar dicas de como os consumidores respondem aos serviços prestados pelo supermercado (ou promoções) e desta forma, auxiliar o gerenciamento do supermercado a estabelecer uma estratégia bem sucedida de retenção de clientes. Entre uma variedade de algoritmos para construir uma árvore de decisões tais como CART (*Classification and Regression Trees* – Árvore de Classificação e Regressão), C5.0, e CHAID (*Chi-squared Automatic Interaction Detection* – Detecção Automática de Interação do Chi-quadrado), selecionou-se o C5.0 devido a sua velocidade, baixa memória requerida, eficiente capacidade de poda, impulso, cruzamento e validação, características que melhoram extremamente a previsão e exatidão.

Frisa ainda o pesquisador que regras “SE \Rightarrow ENTÃO” podem dar ao gerenciamento do supermercado claros perfis de clientes que são alvos de mercado ou ações promocionais. Assim, desenvolveu-se 50 regras diversas, incluindo sub-regras usando o software Clementine 8.0. A exatidão em porcentagem, das regras do “SE \Rightarrow ENTÃO” produzidas por uma árvore de decisões extratoras de informações úteis para formular uma estratégia da retenção, é de 80%.

O autor separou as respostas em apêndice 1, 2, 3, 4 e 5, significando respectivamente o resumo do padrão de compras do consumidor, respeitando o número de itens adquiridos em cada compra, a frequência da visita ao supermercado e o consumidor de um supermercado em particular; o apêndice 2 mostra que o estado civil dos clientes tem algo a ver com seu comportamento de compra solitário ou em grupo; já o 3 ilustra os padrões do relacionamento entre a idade do cliente e a frequência de visita a um supermercado; enquanto o 4 resume os padrões da correlação entre os perfis de cliente (idade, gênero, estado civil, consumidor, volume da compra, frequência de compra) e seu/sua preferência particular e de atributos de serviços. Ainda comentando os critérios de pesquisa de acordo com Min (2006, p.9):

Para investigar mais padrões de relacionamento entre o perfil do cliente e sua dada importância no comportamento do consumidor em determinados atributos de serviço, tentou-se identificar os atributos de serviço que são os mais importantes para os clientes. Para identificar, iniciou-se com causas

determinantes gerais da qualidade do serviço que podem ser relacionadas a qualquer tipo do serviço. Estas causas determinantes incluem confiabilidade, compreensibilidade, competência, acesso, cortesia, comunicação, credibilidade, segurança, compreender o cliente, e tangíveis identificadas por Berry et al. [1985].

Um total de 13 atributos de serviços são identificados e considerados relevantes à qualidade do serviço do supermercado. Como resultado de atributo considerado como o mais importante no que diz respeito à qualidade do serviço do supermercado é a qualidade dos produtos. Dessa maneira, não foi surpresa para o pesquisador que dado supermercado tenha sido escolhido entre os pesquisados como o melhor avaliado quanto à qualidade dos produtos, e sendo este com maior frequência de visitas. Posteriormente, os quatro atributos considerados como os mais importantes foram: limpeza, preço competitivo, variedade de produtos e rapidez no caixa; proximidade da residência e cortesia dos funcionários, além das condições e facilidades de pagamento também foram citadas.

Min (2006, p.10) comenta que pelo fato de que o supermercado lida primeiramente com o alimento que afeta a saúde do cliente e sua família, os clientes tem interesse na condição sanitárias do supermercado, tendo consciência de que os valores gastos com alimentação afetam seu orçamento da renda familiar. Os clientes preferem comprar tudo em uma visita e avaliar uma variedade de produtos, como o quarto atributo importante. Certamente, aproximadamente 15% dos pesquisados indicaram que compraram em média mais de **30 artigos** por a compra.

O pesquisador, entretanto, não notou padrão sistemático correlacionando o perfil do cliente com os quatro atributos considerados como mais importantes, excetuando-se o preço. Isso mostra que não basta apenas saber o perfil do seu cliente. Min (2006) com a pesquisa, também foi capaz de diagnosticar que um cliente solteiro e homem que estivesse no grupo etário entre 30 e 59 anos e comprasse menos de 30 artigos por visita tendem a considerar o serviço de caixa extremamente importante para selecionar o supermercado.

Min (2006) procurou identificar os perfis dos clientes que tendem a consumir em um supermercado. Com isso, o pesquisador demonstrou que cada supermercado pode traçar uma estratégia com base no perfil dos compradores. Por exemplo: enquanto o supermercado *Kroger* procurou criar ações para atrair clientes solitários que não são sensíveis a preços, o supermercado *Meijer* buscou cativar clientes casados dispostos a percorrer uma maior distância para fazer suas compras,

haja visto que, suas lojas são em menor número em bairros residenciais, mas em contrapartida têm mega-lojas em áreas de compras. A tabela 1 mostra os atributos considerados pelos clientes como mais importantes:

Atributos de serviços para supermercados	Grau médio de importância ¹	Rank
Qualidade dos produtos	1.55 (0.88)	1
Limpeza de um supermercado	1.69 (0.87)	2
Preço do competidor	1.72 (0.94)	3
Variedade do produto	1.82 (0.99)	4
Rapidez de caixa	1.88 (0.97)	5
Proximidade à residência	2.03 (0.97)	6
Período de funcionamento da loja	2.19 (1.08)	7
Qualidade de serviços prévios	2.08 (0.94)	8
Boa rotulagem de preço	2.19 (1.00)	9
Cortesia do empregado	2.25 (1.06)	10
Facilidade do pagamento tal como facilidades do cartão de crédito	2.41 (1.25)	11
Departamentos especiais tais como frutos do mar/carnes	2.47 (1.16)	12
Reputação da marca	2.95 (1.07)	13

Tabela 1: Atributos de serviços para supermercado

Nota 1: Os números entre parênteses são desvios padrão.

Fonte: adaptado Min (2006); tabela esta por ordem decrescente

Min (2006) chegou aos seguintes resultados e implicações principais: a lealdade do cliente com relação à sua frequência a determinado mercado e o volume de compras são fatores intimamente ligados. Já a idade e o comportamento têm alguma consequência no que diz respeito ao volume de compras. “Por exemplo, os consumidores de um determinado supermercado tendem a comprar mais artigos à visita do que aqueles que compraram em um número maior de diferentes supermercados”. **O que pode-se concluir deste resultado?** Que a gerência deve imprimir à sua estratégia ações que possam realçar a fidelidade do cliente para aumentar seu rendimento potencial.

Outro resultado relevante é que o cliente que compra pelo menos uma vez por dia, ou seja, o mais freqüente, tende a pagar menos à vista que os menos freqüentes, indicando que, para que a gerência consiga reter e fidelizar este cliente fiel, é necessário pensar em cortesias, ex. disponibilidade maior caixas e auto-serviços) ao cliente que compra um número menor de artigos e pode precisar de serviços mais rápidos em geral.

A combinação estado civil e comportamento de compra resultaram em um padrão sistemático. Isto é, uma pessoa solteira tende a fazer compras sozinha,

resultando em uma decisão única de consumo dentro do supermercado, já as pessoas casadas em contrapartida não, já que costumam fazer as compras em dois. Min (2006, p.13) comenta:

Congruente com o senso comum, um único cliente (solteiro) tende a comprar sozinho. Isto implica que um único cliente é provável fazer só uma decisão de consumo do supermercado. Desde que alguns dos clientes que compram sozinho tendem a preferir o caixa rápido e o pagamento fácil, e localização próxima de uma vizinhança com uma grande porcentagem de clientes solteiros (área universitária) pode melhorar seus serviços para este tipo de cliente aumentando o número de caixas e auto-serviços, ainda aceitando várias formas de pagamento que incluem cartões de crédito e vales. De outro lado, um supermercado que o alvo fosse indivíduos casados (especialmente entre 20, 30 e 40 anos) deve oferecer em sua base uma grande variedade de alimentos e bebidas que atendem às diferentes necessidades dos casais. Também, pode estender seu horário de funcionamento (ex. fins de semana ou mais horas à noite) em que os maridos e as esposas estão disponíveis para as compras.

Um padrão sistemático também foi encontrado entre a idade dos clientes e apreciação pela cortesia dos empregados, ou seja, foi detectado entre os mais jovens - abaixo dos 20 e entre os mais idosos - acima dos 50, uma apreciação maior por atitudes corteses por parte dos empregados do que outras faixas etárias.

A explanação possível para este padrão é que os clientes abaixo de 20 são clientes relativamente inexperientes e podem conseqüentemente precisar o maior auxílio dos empregados do supermercado em encontrar os artigos que querem e validando suas identificações para opções de pagamento sem dinheiro. Similarmente, aqueles clientes que são mais idosos de 50 são mais prováveis sofrer de algum prejuízo físico e precisar subseqüentemente do auxílio dos empregados do supermercado em alcançar artigos nas prateleiras mais altas e em pegar artigos pesados.

O que se pode concluir deste resultado? Que com estas informações em mãos, um supermercado que atenda comunidades de aposentados por exemplo, deve montar estratégias que frisem a cortesia dos empregado como modo de retenção do cliente, investindo pesadamente em treinamento e formação do funcionário.

Outra constatação: os clientes com idade acima de 50 anos são extremamente sensíveis ao preço. Tal descoberta faz sentido já que a maioria destas pessoas depende de aposentadoria e outros recebíveis do governo, e por isso tornam-se mais conscientes do preço do que as demais faixas etárias. Oportunidades de desconto, devem ser estratégias do supermercado que atendem a este tipo de público; por

exemplo: vales, descontos para idosos, bônus fora da compra total, cartões da compra do cliente, entre outros. Em resumo:

[...], um determinado perfil tem um padrão distintivo para consumo. Entretanto, resultados da mineração de dados indicam freqüentemente que uma combinação de características múltiplas do cliente tem a maior influência na seleção do supermercado do cliente e no comportamento do consumo do que um único atributo. Conseqüentemente, em formular uma estratégia bem sucedida da retenção do cliente, a gerência do supermercado deve considerar uma variedade de atributos compreendidos de perfis dos clientes, da freqüência de compra, do volume de compras do mantimento, e da potencialidade do supermercado em oferecer determinadas vantagens (MIN, 2006).

Como última conclusão, Min (2006) detectou que os supermercados têm diferentes perfis de base, isto é, alguns deles direcionam seus serviços para determinados grupos etários, como exemplo destaca que, os supermercados regionais relativamente desconhecidos comparados às grandes redes tais como Pic Pac e Aldi parecem atrair mais freqüentemente uns consumidores mais novos (20 anos) do que outros grupos de idade. Dado sua exposição e recursos limitados, sua estratégia de marketing alveja o nicho de clientes jovens.

A estratégia de retenção de clientes para tais redes seria realçar as cortesias de serviços para essa parte do mercado que deseja atender e atingir. O autor conseguiu identificar muitas regras “Se \Rightarrow então” para padrões de compra de supermercado no decorrer da pesquisa, por exemplo: “SE um cliente vai a um supermercado para comprar mantimento uma vez por dia, A SEGUIR o cliente é provável comprar 10 - 15 artigos por visita”; ou ainda “ SE um cliente vai a um supermercado para a compra de mantimento uma vez por mês E o cliente não é freguês de nenhum supermercado em particular, A SEGUIR o cliente é provável comprar 5 - 10 artigos por visita”.

Não serão citadas todas as regras, porque não é este o foco do presente estudo, contudo, o importante de ser frisado é a importância das técnicas *Data Mining* para a criação de estratégias e de um plano de ações de efetivo sucesso que poderá superar e surpreender a concorrência, fidelizar a clientela e obviamente alavancar as vendas.

2.5 Estudos Anteriores sobre *Data Mining*

Santos et al. (2006), no artigo denominado *Data Mining* aplicado na base de dados das categorias mais vendidas de um supermercado, pretendem responder a seguinte problemática: “qual dia da semana é mais adequado para venda de uma determinada categoria de produtos, considerando-se a proximidade dos feriados e/ou o preço praticado?”

Para tal estudo foram utilizadas técnicas de *Data Mining*, mais especificamente os algoritmos de DM de tarefa de associação (Apriori) em uma amostra dos nove produtos mais vendidos do supermercado. Ademais foram utilizadas outras ferramentas e *softwares* como: “Facilita DM”, *Excel*, *Access* e Bloco de Notas. Optou-se por não revelar a categoria dos produtos analisados.

As abordagens utilizadas para esta pesquisa são tanto de caráter quantitativo (o problema trata de informações numéricas e dados de venda) quanto qualitativo (análise dos resultados com fonte no ambiente natural pesquisado) e também pode ser descritiva quanto à abordagem dos dados. No que se refere aos objetivos, é uma pesquisa de caráter. O trabalho seguiu as etapas do processo KDD⁶ conforme figura 12 a seguir:

⁶ O KDD consiste fundamentalmente na estruturação de banco de dados (BD); na seleção, preparação e pré-processamento dos dados; na transformação, adequação e redução da dimensionalidade dos dados; no processo de DM; e nas análises, assimilações, interpretações e uso do conhecimento extraído do BD. Em geral, num alto nível de abstração, podem-se considerar como sendo três as etapas deste processo: pré-processamento, DM e pós-processamento (SANTOS *et al.*, 2006).

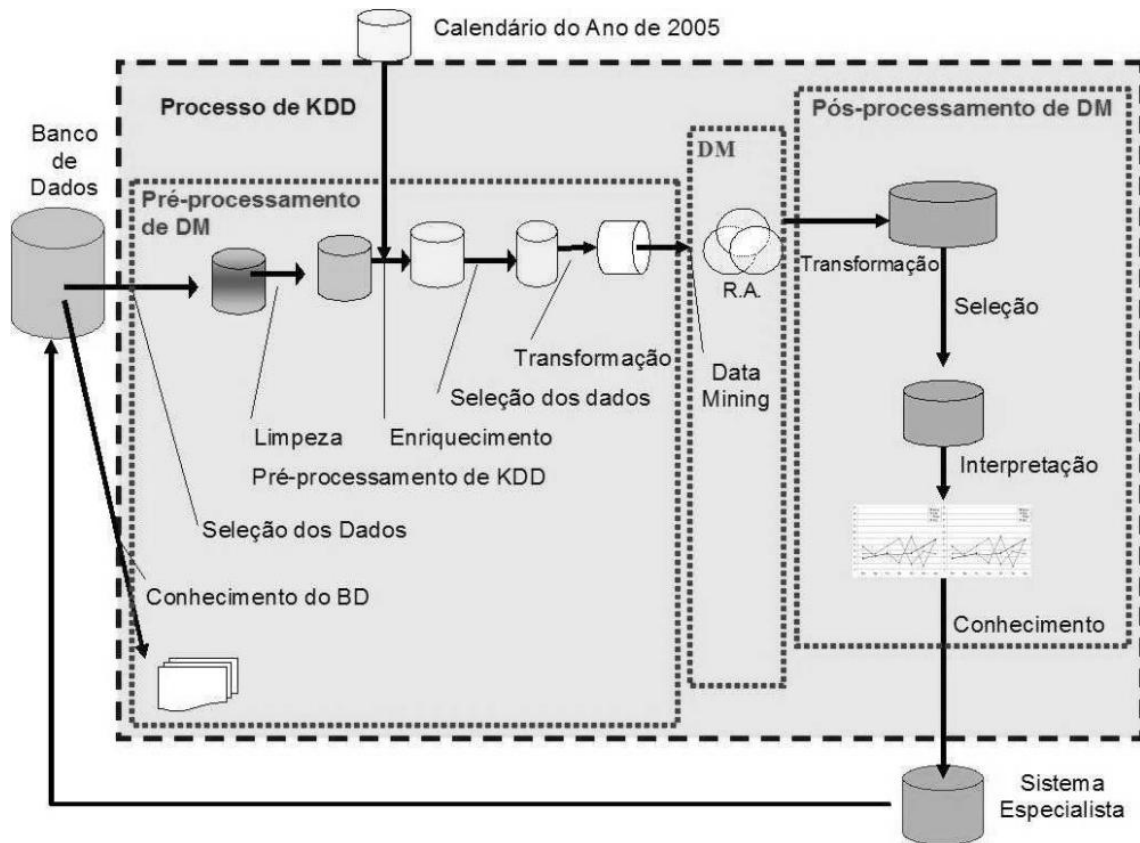


Figura 12 - Etapas do processo de KDD utilizadas no estudo
Fonte: Santos *et al.* (2006).

A partir das estimativas usando as ferramentas de *Data Mining* concluiu-se que a ocorrência dos *tickets* com venda dos produtos das categorias analisadas tende a crescer à medida que se aproxima o fim de semana. Assim, com “grau de confiabilidade superior a 75%”, concluiu-se “que existe uma forte relação entre as categorias vendidas e os dias da semana”.

Os pesquisadores identificaram que as categorias 9 e 2 têm relevância em todos os dias da semana, inclusive no domingo; as categorias 4 e 3 de segunda a sábado, exceto no domingo; a categoria 7 somente de quinta a sábado; as categorias 5 e 6 somente na sexta e sábado; já a categoria 1 apresenta relevância somente no sábado e a categoria 8 não apresentou regras com confiabilidade superior a 75% que associe a mesma com os dias da semana.

Outro estudo sobre o assunto, é a dissertação apresentada por Passari (2003, p.7) denominada *Exploração de Dados Atomizados para previsão de vendas no varejo utilizando redes neurais*. Como problema central o autor tem a seguinte questão: “Como utilizar o processo de mineração de dados para a descoberta de relacionamentos entre produtos e previsão de vendas individual de produtos para apoio à tomada de decisões operacionais no varejo?”.

De acordo com Passari (2003), o “objetivo é explorar um modelo de previsão de demanda para os produtos, visando identificar um composto de marketing adequado (preços, produtos e promoções)”. Como objetivos específicos o autor elenca a exploração de técnicas de redes neurais artificiais para criar modelos de previsão de vendas individualizadas no varejo; bem como, a pesquisa pretendeu desenvolver uma rede neural artificial capaz de detectar relacionamentos entre variáveis que possam impactar no volume de vendas de uma loja ou cadeia de lojas, de forma detalhada (por produto).

Como técnica de pesquisa, o autor usou a modelagem através das redes neurais, a qual é amplamente utilizada em tarefas de classificação e previsão, podendo ser aplicada à geração de conhecimento indireto. No caso específico do artigo de Passari (2003) as redes neurais foram utilizadas para previsão de vendas como já mencionado, “a fim de prever a demanda individualizada de produtos em uma loja de varejo, levando em consideração fatores explicativos da demanda (modelagem causal) e também o relacionamento entre diversos produtos”. A pesquisa seguiu os seguintes passos: obtenção dos dados; seleção e limpeza dos dados; análise do cesto de compras⁷ (envolve: análise do cesto de compras por transação; análise do cesto de compras por cliente; confirmação da análise do cesto de compras); tratamento dos dados para previsão; modelagem com uso de redes neurais e comparação com outras modelagens.

Como conclusão da previsão baseada em redes neurais, o pesquisador salientou que a mesma pode ser considerada adequada para a previsão de demanda de produtos no nível individual, podendo simular preços e estimar nível de estoque por exemplo. “Os resultados sugerem que as redes neurais podem prever as vendas de produtos atomizados no curto prazo com maior precisão, do que as técnicas naive não-ajustada e de regressão linear”. Ademais, as redes neurais podem tratar com facilidade os dados de diferentes produtos e que “informações agregadas sobre um centro de interesse podem auxiliar na previsão, comprovando a existência de relacionamentos entre os produtos, e que tais relacionamentos podem ser captados pelas redes neurais”. Os erros verificados, no entanto, ainda são altos, não sendo possível aproveitar as informações sobre promoções (Passari 2003).

⁷ A análise do cesto de compras é uma técnica de mineração de dados utilizada para encontrar grupos de itens que ocorrem simultaneamente (e em que ordem) em transações (PASSARI, 2003).

Passini e Toledo (2002), em seu artigo Mineração de dados para detecção de fraudes em ligações de água, também descrevem a importância da aplicação da mineração de dados, mais especificamente do *Data Mining* no saneamento básico. O projeto foi desenvolvido com o fim de detectar fraudes em ligações de água, tentando visar as irregularidades e combater as perdas físicas.

Os modelos de análise para mineração de dados visaram traçar o perfil daqueles que fazem ligações clandestinas, visando não só corrigir como também prevenir. Como metodologia, os pesquisadores utilizaram três modelos de mineração de dados; dois deles utilizando o agrupamento neural (para verificar se existia um perfil de fraudes bem definido que pudesse ser qualificado pelo tipo de consumo e categoria, etc) e outro utilizando a classificação por árvore de decisões (para descobrir em que tipo de fraude os consumidores fraudadores se enquadravam). Assim foram separados os que tinham fraudes identificadas dos outros consumidores, sendo documentados os testes realizados em tabelas.

Passini e Toledo (2002) concluíram com a pesquisa, que a técnica de mineração de dados ajuda sim, desde que se tenham dados confiáveis e persistência, já que não se trata de fórmula milagrosa. Ressaltam, no entanto, os pesquisadores que o primeiro modelo foi elaborado por pessoas leigas em estatística e com pouca experiência com a técnica de mineração de dados e dessa forma, os resultados não foram os esperados.

Pinho (2009) apresenta a aplicação de tais mapas com o intuito de identificar perfis de clientes – *clusters* - como estratégia de diferenciação no marketing empresarial. Tais mapas representam uma classe de modelos de redes neurais intituladas mapas auto-organizáveis e podem representar vantagem competitiva, dado seu potencial de aplicações em processos de descoberta de conhecimento.

O artigo de Pinho (2009) foi desenvolvido com a finalidade de *benchmark*, separando clientes de acordo com sua classe de operações na Bovespa em dez possíveis mercados de ação. O pesquisador define mapa auto-organizáveis ou redes Som como “uma classe especial de redes neurais baseadas na aprendizagem competitiva: neurônios na camada de saída competem entre si para ativação ou disparo, e somente um neurônio é ativado a cada apresentação de padrão”, forma-se assim um mapa topográfico de padrões.

Pinho (2009) analisa que a identificação de *clusters* é bastante útil na diferenciação de ações de marketing. Com a prática realizada, os gestores esperam

diminuir a taxa de evasão de clientes a partir do quinto mês de relacionamento. Salienta, no entanto, que alguns pontos de estudo são necessários para melhorar e aperfeiçoar a técnica utilizada.

Giannotti et al. (1999, p.1) apresentam em seu estudo Integração de Dedução e Indução por mineração de dados de vendas de supermercados, o Datasift: “protótipo de um sistema de análise dos dados das vendas de supermercados baseados em técnicas da mineração de dados/*Data Mining*”. Os pesquisadores salientam a importância do *Data Mining* para se descobrir o comportamento do cliente e assim promover ações de marketing. Para esta finalidade, propôs-se uma arquitetura que integra as capacidades dedutivas de uma língua baseada na lógica da base de dados, LDL++, com a capacidade indutiva da diversidade de algoritmos e de ferramentas da mineração de dados.

O objetivo da pesquisa de Giannotti et al. (1999, p.3) foi ilustrar a expressividade alcançada a partir da aplicação da análise integrada do sistema de segmentação de mercado, mais especificamente como uma “integração apropriada do raciocínio dedutivo, tal como aquela suportada por línguas da base de dados da lógica, e o raciocínio indutivo, tal como aquele suportado por regras de associação, fornecem uma solução viável” aos problemas na análise da cesta do mercado.

Os pesquisadores concluíram que o sistema revelou-se ferramenta eficaz para os analistas de mercado; ademais, de acordo com os mesmos autores, “experimenta-se que um coordenador perito em conhecimento pode geralmente traduzir rapidamente as perguntas das regras dedutivas/indutivas”. Um dos problemas encontrados no protótipo foi a expansibilidade na interface do utilizador, já que a cada nova análise a interface precisa ser modificada.

3 METODOLOGIA

3.1 Pesquisa Descritiva

A pesquisa descritiva é definida por Vieira (2002, p.5) como aquela cujo “interesse é descobrir e observar os fenômenos, sem, contudo causar interferência no intuito de modificá-la”. Pode aproximar-se das pesquisas experimentais, à medida em que se interessa pelas relações entre as variáveis. Ainda, a pesquisa descritiva, de acordo com o mesmo autor, “expõe as características de determinada população ou de determinado fenômeno, mas não tem o compromisso de explicar os fenômenos que descreve, embora sirva de base para tal explicação”.

Os estudos comumente utilizados neste tipo de pesquisa são o longitudinal, cujas informações vão sendo coletadas ao longo do tempo; e o transversal, que diz respeito ao presente estudo, cujas informações são coletadas uma vez no tempo. Esta investigação usa a pesquisa descritiva para salientar as associações de compras em supermercado, descrevendo-as.

3.2. Banco de Dados

Dados: Os dados nesta pesquisa são do tipo secundário. Dados secundários são definidos por Gomes (2005, p.20) como aqueles já existentes nas mais diversas fontes, sejam elas, internet, jornais, revistas, dentre outras. Os dados secundários podem ser coletados a partir de um banco de dados interno ou externo a organização, como pesquisa em sites governamentais, como IBGE- Instituto Brasileiro de Geografia e Estatística e IPEA- Instituto de Pesquisa Econômica por exemplo; ou publicações empresariais e setoriais; serviços de pesquisas como o IBOPE; pesquisas de potencial de consumo e grupos profissionais como associações ou instituições de classe.

Os dados **secundários** podem ser coletados também, como foi o foco do presente estudo, de um banco de dados interno. Para Gomes (2005, p.20), uma empresa possui informações que, se reunidas, relacionadas e bem utilizadas podem se transformar em instrumentos importantes para a tomada de decisões, como exemplo é possível se extrair informações do sistema de pedidos da empresa e de seus vendedores, como: detalhamento do número de vendas efetuadas; Características dos clientes; Informações sobre a concorrência, tais como: preços,

formas de pagamento, novos produtos.

No caso específico do presente estudo, tais informações foram coletadas do banco interno do supermercado e tratados através da ferramenta *Data Mining*.

3.3 Período de Tempo Longitudinal

Pesquisa com investigação longitudinal é um estudo que se realiza ao longo de um período de tempo. Os dados são colhidos primeiramente no início do estudo, e podem então ser recolhidos outras vezes ao longo de todo o cumprimento do estudo. Alguns estudos longitudinais podem ocorrer durante um curto período de tempo, como de alguns dias, enquanto outros podem perdurar um período de décadas.

Realizou-se um estudo descritivo a partir do levantamento do universo de registros das vendas correspondente ao período de **Janeiro de 2009 até Dezembro de 2009**, totalizando **304.522** transações no período analisado. Para este trabalho foi considerada venda, toda operação iniciada no caixa com a abertura de um cupom fiscal e finalizada pelo mesmo.

3.4 Amostra

Escolha da amostra: O supermercado Franzoni & Cia Ltda, situado em Quedas do Iguaçu, foi escolhido em virtude de ser o maior supermercado da cidade tanto no que tange faturamento (não autorizado a divulgar). Tem como concorrentes diretos na cidade cinco supermercados e diversas mercearias e mini-mercados. O supermercado Franzoni & Cia Ltda é conceituado na região sudoeste do estado do Paraná como inovador dentro do seu segmento e muito ativo na comunidade.

Foram analisados alguns produtos e o comportamento dos consumidores frente a estes produtos. Posteriormente, a partir do tratamento dos dados do *Data Mining* foram organizados combos de produtos, que foram colocados à venda por tempo determinado. Depois deste tempo pré-determinado foi analisado se os combos formados surtiram resultados em termos de venda.

3.5 Análise de Dados e Regras de Associação

A tarefa de associação (HAN; KAMBER, 2001) utilizada neste trabalho tem por objetivo analisar padrões em um conjunto de dados que contêm itens que estão relacionados à ocorrência de outros itens nas operações de vendas em supermercado.

É importante considerar os conceitos que foram comumente utilizados durante a apresentação dos resultados da mineração de dados. Agrawal et al. (1993), foram os primeiros a introduzir as regras/conceitos de associação. Elas são:

- Regras de Associação: são resultados da mineração de dados. Uma das principais funções é representar padrões de compra na qual existem dois ou mais produtos associados. Toda regra tem seu “espelho”, isso é a relação inversa. No exemplo, na regra “ $A \Rightarrow B$ ”, a aquisição do produto “A” implica na compra do produto “B”. Logo, também haverá a regra “ $B \Rightarrow A$ ”.
- Suporte: é o número de transações em que a associação ocorre no período analisado. Por exemplo: a quantidade de vezes que ocorreu a transação abaixo foi de 1.113 vezes.

Pão de forma doce \Rightarrow Pão de sal = - 1.113

- Confiança: Gonçalves (2010), indica que a medida confiança representa dentre as transações que possuem os itens de A, a porcentagem de transações que possuem também os itens de B, indicando a validade da regra. No exemplo abaixo ficou em 35%.

Pão de forma doce \Rightarrow Pão de sal = - 1.113 - 35%

- Transacional: A base de dados transacionais contém todas as informações operacionais da empresa.

Seja $I = \{I_1, I_2, \dots, I_n\}$ um conjunto de itens. Seja D uma base de dados de transações, em que cada transação T é formada por um conjunto de itens onde $T \subseteq I$. Cada transação possui um identificador chamado TID. Uma regra de associação é uma implicação da forma $A \Rightarrow B$, onde A e B podem

ser conjuntos compostos por um ou mais itens, $A \subset I$, $B \subset I$, e $A \cap B = \emptyset$. A é chamado de antecedente da regra e B é chamado de conseqüente.

Dada uma regra $A \Rightarrow B$, a sua medida de suporte representa a porcentagem de transações da base de dados que contêm os itens de A e B, indicando a relevância da mesma. Já a sua medida de confiança representa, dentre as transações que possuem os itens de A, a porcentagem de transações que possuem também os itens de B, indicando a validade da regra. O problema da mineração de regras de associação, conforme definido originalmente em 1993, consiste em encontrar todas as regras de associação que possuam suporte e confiança maiores ou iguais, respectivamente, a um suporte mínimo (SupMin) e uma confiança mínima (ConfMin), especificados pelo usuário (GONÇALVES, 2010, p. 1).

Em mineração de dados, aprendizagem de regra de associação é um método popular e bastante pesquisado para descobrir relações interessantes entre variáveis em grandes bases de dados. Piatetsky-Shapiro (1991) descreve-a apresentando fortes regras descobertas em base de dados usando diferentes medidas de interesse. Baseado no conceito de regras fortes, Agrawal et al. (1993) introduziram regras de associação para descobrir regularidades entre produtos em transações de grande escala nos dados registrados por ponto-de-venda (POS) em sistemas de supermercados.

Por exemplo, a regra $\{\text{cebola, batatas}\} \Rightarrow \{\text{carne}\}$ descoberta nos dados de vendas de um supermercado indicaria que se um cliente comprasse cebolas e batatas juntas, seria provável que ele ou ela também comprasse carne. Tais informações poderiam ser usadas como base para decisões sobre atividades de marketing tal como, por exemplo, fixação de preços promocionais ou disposição de produtos. Além do exemplo anteriormente mencionado, regras de associação de análise de cesto de mercado são empregados hoje em muitas áreas de aplicação, incluindo mineração de uso de web, detecção de intrusão e bioinformática.

No intuito de explicar como funciona este processo, apresenta-se no quadro 7 um exemplo em uma pequena base de dados, a qual armazena as compras feitas por clientes de um supermercado fictício

1. Biscoito, **cerveja**, chá, **salaminho**.
2. Cerveja, couve, lingüiça, pão, queijo.
3. Café, brócolis, couve, pão.
4. Brócolis, café, **cerveja**, couve, pão, **salaminho**.
5. Brócolis, café, couve, pão, refrigerante.
6. Couve, lingüiça.

Quadro 7: Exemplo de base de dados com compras realizadas
 Fonte: Gonçalves (2010, p.1).

Pode-se observar que cada um dos registros da base de dados armazena a relação de produtos que foram adquiridos por um cliente específico. Como menciona Gonçalves (2010, p.1):

Um exemplo de regra de associação que poderia ser minerada nesta base de dados, através da utilização de uma ferramenta de *Data Mining*, é dado por: {*cerveja*} \Rightarrow {*salaminho*}. Note que duas das seis transações que compõem a base contêm os produtos {*cerveja*} e {*salaminho*}. Desta maneira, o suporte da regra {*cerveja*} \Rightarrow {*salaminho*} pode ser calculado da seguinte forma: $2 \div 6 = 33,33\%$. Observe agora que na base de dados, existem duas transações que contêm os produtos {*cerveja*} e {*salaminho*} juntos e três transações que contêm o produto {*cerveja*}. A confiança da regra {*cerveja*} \Rightarrow {*salaminho*} pode então ser calculada da seguinte maneira: $2 \div 3 = 66,67\%$. Este valor indica que 66,67% dos consumidores que compraram {*cerveja*} também compraram {*salaminho*}.

De acordo com Lovin (2008, p.1), determinam-se os *itemsets* (grupos de itens) freqüentes baseados num suporte predefinido. Tem-se todos os itens que freqüentemente são vendidos juntos. Essa informação por si só é válida, mas necessita-se de mais. Precisa-se saber qual é o piloto dentro de um *itemset* freqüente.

Por exemplo, um agente de seguros diria que um cliente que compra um Benefício de Renda de Deficiência, na maior parte do tempo, também compraria um Prêmio para Benefício de Deficiência. Em outras palavras, o consumidor que compra um Benefício de Morte Acidental, também compraria um Benefício Invalidez. Quais fatos desconhecidos levam o cliente a comprar itens que aparentemente não tem nenhuma relação? É neste momento que encontrar regras de associação torna-se útil.

O estudo em questão foi realizado conforme esquema descrito a seguir:

Pão de forma doce	=>	Pão de sal	=	1.113	-	35%
Pão de sal	=>	Pão de forma doce	=	1.113	-	1%

A compra associada dos produtos Pão de sal e Pão de forma doce na mesma transação aconteceu 1.113 vezes (suporte) no período analisado. Sob a perspectiva de compra do Pão de sal em função do Pão de forma doce, as chances são de 35%. Ou seja, a cada 100 transações envolvendo Pão de forma doce, 35 provavelmente conterão também Pão de sal.

Por outro lado, quando se analisa as transações em função do produto Pão de sal, a chance de haver a venda conjunta com Pão de forma doce é de 1%. A redução da probabilidade justifica-se pela grande procura por Pão de sal e a conseqüente venda associada a outros diversos produtos, como leite, presunto e café.

Assim, a regra de associação possibilita através da combinação de itens adquiridos pelo consumidor e analisada pela ferramenta *Data Mining* a entender e considerar as variáveis suporte e confiança para que possamos realizar os cálculos necessários neste trabalho. A forma como foi extraído os dados da empresa pesquisada esta representada no quadro 8:

```

* Código (espaço) (aspas) NOME DO PRODUTO COMPLETO (aspas)
7898422747916 "SHAMPO DOVE 250 ML "
7896054905414 "PIPOCA MICRO 100 G "
7898043362703 "CAD XADREZ 96 F "
7890026024981 "RELOGIO MESA "
7896020156888 "SABONETE SNOOPY 80G "
7891354044108 "TOALHA INFANTIL "
7891155031727 "COPO ESMILINGUIDO "
7896275960896 1.00 7891000053508 1.00 7891095100293 1.00 #
0000000004541 1.00 7898196701084 1.00 7896054711114 1.00
7896098900116 1.00 7896321007223 2.00 #
7891991006200 1.00 #
7892840200145 1.00 00000000000192 2.00 00000000004824 1.00
7891700205030 1.00 7896014184996 1.00 7898180570191 1.00 #
7896232800470 1.00 7896114100049 1.00 7894904009862 1.00
7896102517217 1.00 #
7891038160407 1.00 7891038009300 1.00 #
7896798500524 12.01 7898007821093 1.00 7898422745226 1.00
7891055337349 1.00 7891080404887 1.00 #
7898422754280 1.00 7501001156176 1.00 7891088191253 1.00 #
7896026800068 1.00 7896040703963 1.00 7622300261795 1.00
7891099691186 1.00 00000000000192 2.00 0000078912366 2.00 #
7891038009300 1.00 7896104721100 2.00 7894900011593 1.00
00000000002097 1.00 7896526100019 2.00 00000000007405 2.00
7896389300045 1.00 7891095006243 1.00 7891095007240 1.00
7891095002603 1.00 00000000002134 1.00 #
7891040128099 1.00 7898910365066 2.00 7891095010233 1.00
7896283000058 1.00 7891055805602 1.00 7896486201573 1.00
7891035250200 1.00 7896495000129 1.00 7896026800075 1.00
4005808213351 1.00 7890704806502 1.00 7897590100172 1.00
00000000005104 1.00 00000000005241 1.00 7897664185111 1.00
7896017455505 2.00 7891024194102 1.00 00000000002097 1.00
7898187730123 1.00 7894900010015 1.00 7896104867389 1.00
7891051028791 1.00 #
* cada linha é uma compra com 1 ou N produtos
* a ordem é a seguinte: código do produto (espaço) quantidade vendida (espaço) ...
# para finalizar a linha
* quantidade pode ser fracionada, como em peso (0,120)

```

Quadro 8: Exemplo das informações extraídas do supermercado e seus códigos.
Fonte: Pesquisa.

Importante também frisar que durante o mês de Maio/2010 e Junho/2010 – tendo como período de amostra o intervalo dos dias 01 a 31 – foram sugeridos quatro **combos** no supermercado pesquisado, para que se possa comparar a vendagem dos produtos no mesmo período do ano passado e validar o que foi sugerido. Os combos foram evidenciados no supermercado, através de uma disposição conhecida como ponta de gôndola.

3.6 Histórico da Empresa Pesquisada

O Sr. Adelino Franzoni e a Sr^a Michelina Dilina Losso Franzoni iniciaram as atividades comerciais no segmento de roupas e calçados no dia 10 de janeiro de 1973, à Rua Barão do Rio Branco, 1014 na cidade de Guarapuava, estado do Paraná, do Comercial Franzoni de Roupas Feitas e Calçados Ltda, onde mais tarde seria criada a empresa **Franzoni & Cia Ltda**.

No dia 05 de dezembro de 1974 a sede da Empresa foi transferida para a cidade de Quedas do Iguaçu - PR, e acrescido em suas atividades o comércio de secos e molhados em geral, tendo um quadro de dezesseis funcionários. No ano de 1977 com o desenvolvimento da cidade, especialmente das indústrias madeireiras, houve um incremento nos negócios, passando a atuar com um supermercado, localizado na Rua Acácia, 148 nesta cidade.

Em 1986 as atividades foram separadas, ficando na Rua Acácia a Loja de confecções e Calçados e na Rua Jacarandá era inaugurado um novo supermercado com um espaço físico mais adequado para época. Já neste período trabalhavam nas empresas quarenta e cinco funcionários.

No dia 29 de novembro de 2000, a empresa transferiu as atividades para a rua Juazeiro, 1429 em Quedas do Iguaçu, PR, onde permanece até os dias atuais, para melhorar o atendimento e disponibilizar maior conforto aos clientes, fornecedores e funcionários. E é neste local que a pesquisa exploratória está sendo desenvolvida. Possui uma área construída de 4400 m², sendo de 2400 m² de área de venda, contando com 14 check-outs, estacionamento coberto, ainda com mais de vinte seções, sendo as mais importantes: alimentos, limpeza, higiene, utilidades, papelaria, frios e congelados, hortifrutigranjeiros, bebidas, açougue, lanchonete, industrializado, eletrodoméstico, panificadora, brinquedos, doces, *pet shop* e outros.

Dentre os objetivos principais do supermercado, está em oferecer maior poder de escolha, produtos e serviços com qualidade superior, tornando possível proporcionar às pessoas uma melhor qualidade de vida.

O Supermercado conta com **uma** filial desde 2006 na mesma cidade, localizado no bairro Luzitani, onde conta com mais aproximadamente 30 colaboradores, 8 check-outs, com estrutura em divisão de setores similar à matriz, trabalhando com um número de itens inferior a matriz, com sede própria, o faturamento anual não foi revelado.

3.7 Varejista

A seguir são apresentadas as fotos do varejista. O supermercado analisado foi o Supermercado Franzoni & Cia Ltda. A foto 1, lado direito, busca demonstrar a exposição das bebidas, realçando a extensão na parte lateral do supermercado, a foto 2 mostra a parte dos hortifrutigranjeiros, a foto 3 evidencia os 14 *check-outs* que facilitam a vida dos consumidores, a foto 4 realça o estacionamento coberto, com segurança monitorada a disposição dos consumidores, e a foto 5, mostra a lanchonete instalada no supermercado.



Foto 1: Supermercado Franzoni & Cia Ltda



Foto 2: Supermercado Franzoni & Cia Ltda - Setor hortifrutigranjeiros



Foto 3: Supermercado Franzoni & Cia Ltda – 14 Check-outs



Foto 4: Supermercado Franzoni & Cia Ltda – Estacionamento com capacidade 60 carros



Foto 5: Supermercado Franzoni & Cia Ltda - Lanchonete

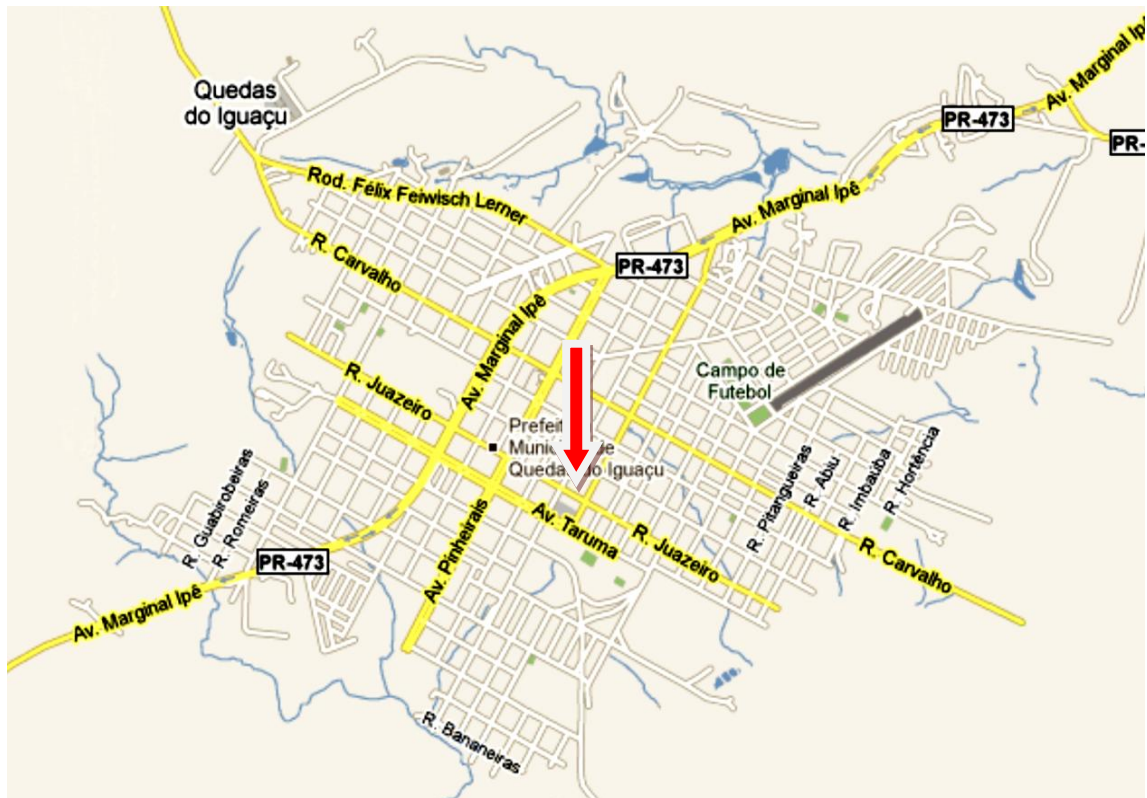


Foto 6: Mapa da cidade de Quedas do Iguaçu – indicando a localização do Supermercado.

4 ANÁLISE DOS DADOS

4.1 Total de Produtos no Supermercado – Análise Anual

Este tópico apresenta os resultados dos dados de mineração do supermercado.

De acordo com a base de dados obtida foi possível segmentá-la em três grandes grupos – **anual**, **bimestral** e **segmentos de produtos**. Assim sendo, foram pesquisadas as regras de associação mais frequentes no ano. A fim de aumentar a qualidade dos resultados e permitir análises comparativas também foram consideradas as bases separadamente em bimestres e depois bases individuais relevantes de alguns produtos selecionados.

É importante considerar a definição de termos e conceitos que serão comumente utilizados durante a apresentação dos resultados da mineração de dados. **Regras de Associação** são resultados da mineração de dados. Portanto, uma das principais funções é representar padrões de compra na qual existem dois ou mais produtos associados. Toda regra tem seu “espelho”, isso é a relação inversa. No exemplo “A => B”, a aquisição do produto “A” implica na compra do produto “B”. Logo, também haverá a regra “B => A”.

Portanto, serão apresentadas as 5 regras mais comuns a cada bimestre, 15 regras mais comuns ao ano e produtos individuais relevantes. A classificação das associações foi estabelecida sob os seguintes critérios:

- a) **Maior suporte - número de vezes da ocorrência**
- b) **Maior confiança - probabilidade de acerto da associação(%)**

Discutir-se-á agora o período anual de Janeiro até Dezembro de 2009. O ponto de corte (critério) foi estabelecido através de compras com suporte acima de 2.000, para que fosse realizada apenas análises de produtos considerados relevantes para o supermercado e confiança de 0 até 100%. Assim, busca-se as maiores ocorrências de frequência de compra em comportamento do consumidor.

Período: Jan - Dez					
Item	Lista 1		Lista 2	Suporte	Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG	3.818	28%
	BANANA CATURA KG	=>	BATATA MONALIZA KG	3.818	22%
2	TOMATE L.VIDA KG	=>	BATATA MONALIZA KG	2.921	29%
	BATATA MONALIZA KG	=>	TOMATE L.VIDA KG	2.921	22%
3	TOMATE L.VIDA KG	=>	BANANA CATURA KG	2.866	28%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG	2.866	16%
4	ALFACE CRESPI UN	=>	BANANA CATURA KG	2.815	24%
	BANANA CATURA KG	=>	ALFACE CRESPI UN	2.815	16%
5	ERVA VERDE LAND 1 KG	=>	BATATA MONALIZA KG	2.559	32%
	BATATA MONALIZA KG	=>	ERVA VERDE LAND 1 KG	2.559	19%
6	ALFACE CRESPI UN	=>	BATATA MONALIZA KG	2.470	21%
	BATATA MONALIZA KG	=>	ALFACE CRESPI UN	2.470	18%
7	ERVA VERDE LAND 1 KG	=>	BANANA CATURA KG	2.379	30%
	BANANA CATURA KG	=>	ERVA VERDE LAND 1 KG	2.379	13%
8	SAL MOC 1 KG	=>	BATATA MONALIZA KG	2.362	35%
	BATATA MONALIZA KG	=>	SAL MOC 1 KG	2.362	17%
9	LARANJA KG	=>	BANANA CATURA KG	2.300	29%
	BANANA CATURA KG	=>	LARANJA KG	2.300	13%
10	CEBOLA NAC. KG	=>	BATATA MONALIZA KG	2.154	32%
	BATATA MONALIZA KG	=>	CEBOLA NAC. KG	2.154	16%
11	REPOLHO VERDE KG	=>	BANANA CATURA KG	2.128	30%
	BANANA CATURA KG	=>	REPOLHO VERDE KG	2.128	12%
12	REPOLHO VERDE KG	=>	BATATA MONALIZA KG	2.116	30%
	BATATA MONALIZA KG	=>	REPOLHO VERDE KG	2.116	16%
13	TOMATE L.VIDA KG	=>	ALFACE CRESPI UN	2.109	20%
	ALFACE CRESPI UN	=>	TOMATE L.VIDA KG	2.109	18%
14	SAL MOC 1 KG	=>	BANANA CATURA KG	2.100	31%
	BANANA CATURA KG	=>	SAL MOC 1 KG	2.100	12%
15	CEBOLA NAC. KG	=>	BANANA CATURA KG	2.066	31%
	BANANA CATURA KG	=>	CEBOLA NAC. KG	2.066	12%

Tabela 2: Principais associações no período de Janeiro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

De acordo com a **Tabela 2**, item 1, nota-se que a maior associação encontrada foi entre a batata monaliza kg e banana catura kg, a qual aconteceu **3.818** vezes (suporte) no período analisado. Em outras palavras, 3.818 vezes divididos por 12 meses temos 318.16 operações mensais, ainda divididos por 25 dias em média que o supermercado em estudo trabalha por mês, esta compra ocorreu em média **12,72** vezes ao dia. Suporte é o número de transações em que a associação ocorre no período analisado.

A confiança da associação batata monaliza kg e banana catura kg foi de **28%**, indicando que a cada cem transações com produtos batata monaliza kg é possível que o consumidor adquira 28 vezes a banana catura kg. Confiança é o valor percentual que representa a possibilidade de a transação associada ocorrer em função de cada produto em uma regra de associação.

Avaliando a regra inversa, a compra associada dos produtos banana catura kg e batata monaliza kg possui o mesmo suporte **3.818** vezes, desta forma, a confiança da associação banana catura kg e batata monaliza kg foi de 22%, ou seja, menor do que a anterior. Esta situação pode ser percebida pelos grupos de interesse de compras que podem ser estabelecidos com a aquisição do produto banana catura, como, por exemplo, o consumo do produto isolado, produtos para sobremesa, e outros.

Na **Tabela 2**, item 5, observando a compra associada dos produtos erva verdeland 1kg e batata monaliza kg no período analisado, somaram **2.559** vezes (suporte). O supermercado pesquisado trabalha na média 25 dias por mês, assim, 2.559 vezes divididos por 25 dias, esta compra ocorreu em média **8,53** vezes ao dia.

A confiança da associação erva verdeland 1kg e batata monaliza kg foi de 32%, assim, de cada 100 transações com produtos erva verdeland 1kg é provável que o consumidor adquira 32 vezes a batata monaliza kg.

Verificando a regra de forma inversa, a compra associada dos produtos erva verdeland 1kg e batata monaliza kg apontam suportes iguais em 2.559 vezes, entretanto, a confiança da erva verdeland 1kg e batata monaliza kg foi de 19%, inferior a regra $A \Rightarrow B$. Os números encontrados podem estar alicerçados principalmente pela migração “gaúcha” nesta região e demonstra através do hábito de consumo da erva mate os números descobertos, podendo inclusive sugerir alteração no posicionamento de gôndolas em lugares estratégicos, associando com outros produtos na tentativa de alavancar vendas.

No item 8, da **Tabela 2**, verificando a compra associada dos produtos sal moc 1kg e batata monaliza kg durante o período pesquisado, esta transação ocorreu 2.362 vezes (suporte). Realizando a divisão de suporte (2.362) pela média de dias trabalhados (25) no mês pelo supermercado pesquisado, esta compra aconteceu em média 7,87 vezes ao dia.

A confiança da associação sal moc 1kg e batata monaliza kg foi de 35%, desta forma, existe a probabilidade de que a cada cem operações com produtos sal moc 1kg o consumidor possa vir adquirir em 35 delas batata monaliza kg.

Verificando a regra por outro lado, a compra associada dos produtos batata monaliza kg e sal moc 1kg identificados pelo suporte de 2.362 vezes. Todavia, a confiança do sal moc 1kg e batata monaliza kg foi de 17%, percentual inferior ao encontrado anteriormente. Isso é possível verificar através do grau de utilização destes produtos no consumo diário. O sal esta presente em uma grande quantidade de alimentos consumidos diariamente.

De acordo com a **Tabela 2**, item 13, analisando a compra associada dos produtos tomate l.vida kg e alface crespa un durante o período estudado, a mesma transação ocorreu 2.109 vezes (suporte). Efetuando a divisão do suporte pela quantidade de dias (25) em média que o supermercado em estudo trabalha, esta compra ocorreu em média **7,03** vezes ao dia.

A confiança da associação tomate l.vida kg e alface crespa un foi de **20%**, indicando que a cada cem transações com produtos tomate l.vida kg existe a possibilidade de que o consumidor adquira em 20 vezes a alface crespa un.

Avaliando a regra inversa, a compra associada dos produtos tomate l.vida kg e alface crespa un possui suporte idêntico de 2.109 vezes, contudo, a confiança da alface crespa e tomate l.vida kg un foi de 18%, menor que a anterior. Estes números podem refletir a maior possibilidade de combinações para utilização do produto tomate l.vida kg e mesmo no consumo individual, principalmente na culinária.

Na tabela anual, o ponto de corte (critério) estabelecido foi para compras com suporte acima de 2.000 e confiança de 0 até 100%. Agora, o ponto de corte (critério) estabelecido é com suporte maior que 500, confiança de 0 até 100% para as tabelas bimestrais. Assim, busca-se as maiores frequências de compra no comportamento do consumidor.

4.2 Total de Produtos no Supermercado – Análise Bimestral

Através da análise bimestral, busca-se demonstrar o comportamento das associações de compras em períodos menores, possibilitando uma maior compreensão dos dados descobertos no período anual e verificando a correlação dos resultados entre os períodos.

Discutir-se-á agora o período bimestral de Janeiro até Fevereiro de 2009. De acordo com a **Tabela 3**, item 1, nota-se que a maior associação foi entre a batata monaliza kg e banana catura kg (714).

Período: Jan - Fev						
#	Lista 1		Lista 2		Suporte	Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG		714	29%
	BANANA CATURA KG	=>	BATATA MONALIZA KG		714	22%
2	ALFACE CRESPA UN	=>	BANANA CATURA KG		565	24%
	BANANA CATURA KG	=>	ALFACE CRESPA UN		565	17%
3	TOMATE L.VIDA KG	=>	BANANA CATURA KG		537	28%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG		537	16%
4	LARANJA KG	=>	BANANA CATURA KG		536	30%
	BANANA CATURA KG	=>	LARANJA KG		536	16%
5	ALFACE CRESPA UN	=>	BATATA MONALIZA KG		522	22%
	BATATA MONALIZA KG	=>	ALFACE CRESPA UN		522	21%

Tabela 3: Associações no período de Janeiro-fevereiro de 2009 com suporte relevantes
Fonte: Pesquisa.

A compra associada dos produtos batata monaliza kg e banana catura kg na mesma operação aconteceu **714** vezes (suporte) no período analisado. Em outras palavras, 714 vezes por 25 dias em média que o supermercado em estudo trabalha no mês, esta compra ocorreu em média **14,28** vezes ao dia.

A confiança da batata monaliza kg e banana catura kg foi de **29%**, assim, a cada 100 transações com produto batata monaliza é possível que o consumidor adquira 29 vezes a banana catura kg.

Avaliando a regra de forma inversa, a compra associada dos produtos banana catura kg e batata monaliza kg possui exatamente o suporte de 714 vezes, entretanto, a confiança da banana catura kg e batata monaliza kg foi de 22%. Cotejando percentuais da regra $A \Rightarrow B$ e $B \Rightarrow A$, fica evidenciado a importância do produto batata monaliza kg, podendo sugerir aos gestores responsáveis por compras, preços, promoções, exposição e reposição do produto uma atenção especial.

Com base na **Tabela 4**, item 1, período de Março até Abril de 2009, a compra associada dos produtos batata monaliza kg e banana catura kg na mesma transação ocorreram **657** vezes (suporte) no período indicado. Desta forma, 657 vezes por 25 dias em média que o supermercado em estudo trabalha no mês, esta compra ocorreu em média **13,14** vezes ao dia. A confiança da associação batata monaliza kg e a banana catura kg foi de 28%.

Avaliando a regra $B \Rightarrow A$, a compra associada dos produtos banana catura kg e batata monaliza kg possui o mesmo suporte 657 vezes, desta forma, a confiança da banana catura kg e batata monaliza kg foi de 22%, comportamento similar ao identificado no período de Janeiro até Fevereiro e no período anual, reforçando a posição destes produtos dentro do mix hortifrutigranjeiro.

Período: Mar - Abr				
#	Lista 1		Lista 2	Suporte Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG	657 28%
	BANANA CATURA KG	=>	BATATA MONALIZA KG	657 22%
2	ALFACE CRESPA UN	=>	BANANA CATURA KG	573 25%
	BANANA CATURA KG	=>	ALFACE CRESPA UN	573 19%
3	TOMATE L.VIDA KG	=>	BATATA MONALIZA KG	563 28%
	BATATA MONALIZA KG	=>	TOMATE L.VIDA KG	563 24%
4	TOMATE L.VIDA KG	=>	BANANA CATURA KG	562 28%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG	562 19%
5	ALFACE CRESPA UN	=>	BATATA MONALIZA KG	512 22%
	BATATA MONALIZA KG	=>	ALFACE CRESPA UN	512 22%

Tabela 4: Associações no período bimestral de Março-Abril de 2009 com suporte relevantes
Fonte: Pesquisa.

Discutir-se-á agora o período bimestral de Maio até Junho de 2009. De acordo com a **Tabela 5**, item 1, percebe-se que a maior associação foi entre a batata monaliza kg e banana catura kg.

A compra associada dos produtos batata monaliza kg e banana catura kg na mesma operação foi de **503** vezes (suporte) no período apurado. Assim, 503 vezes por 25 dias em média que o supermercado em estudo trabalha no mês, esta

compra ocorreu em média 10,06 vezes ao dia. A confiança da associação batata monaliza kg e a banana catura kg foi de 27%.

Avaliando a regra inversa, a compra associada dos produtos banana catura kg e batata monaliza kg possui o mesmo suporte 503 vezes, contudo, a confiança da batata monaliza kg e banana catura kg foi de 21%.

Período: Mai - Jun				
#	Lista 1		Lista 2	Suporte Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG	503 27%
	BANANA CATURA KG	=>	BATATA MONALIZA KG	503 21%
2	ALFACE CRESPA UN	=>	BANANA CATURA KG	390 25%
	BANANA CATURA KG	=>	ALFACE CRESPA UN	390 16%
3	TOMATE L.VIDA KG	=>	BATATA MONALIZA KG	383 27%
	BATATA MONALIZA KG	=>	TOMATE L.VIDA KG	383 21%
4	TOMATE L.VIDA KG	=>	BANANA CATURA KG	379 26%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG	379 16%
5	SAL MOC 1 KG	=>	BANANA CATURA KG	354 32%
	BANANA CATURA KG	=>	SAL MOC 1 KG	354 15%

Tabela 5: Associações no período bimestral de Maio-Junho de 2009 com suporte relevantes
Fonte: Pesquisa.

Analisando o período bimestral de Julho até Agosto de 2009, de acordo com a **Tabela 6**, item 1, fica evidenciado que a maior associação foi entre batata monaliza kg e banana catura kg. A compra associada dos produtos batata monaliza kg e banana catura kg somaram no período 610 vezes (suporte). Assim, 610 vezes divididos por 25 dias em média que o supermercado em estudo trabalha no mês, esta compra ocorreu em média **12,20** vezes ao dia. Indicando que confiança da associação batata monaliza e a banana catura foi de **29%**.

A regra inversa demonstra que a compra associada dos produtos banana catura kg e batata monaliza kg possui o mesmo suporte 610 vezes, todavia, a confiança da batata monaliza kg e banana catura kg foi de 22%.

De acordo com a **Tabela 7**, item 1, do período bimestral de Setembro até Outubro de 2009, observou-se que a maior associação foi entre a batata monaliza kg e banana catura kg.

A compra associada dos produtos batata monaliza kg e banana catura kg na mesma operação aconteceu 728 vezes (suporte) no período analisado. Avaliando a quantidade de vezes (728) suporte divididos pela quantidade de dias médios trabalhados no mês (25), esta compra ocorreu em média **14,56** vezes ao dia. Gerando uma confiança da associação batata monaliza e banana catura de 30%.

Na avaliação da regra inversa, a compra associada dos produtos banana catura kg e batata monaliza kg geraram suporte igual de 728 vezes, não obstante, a confiança da batata monaliza kg e banana catura kg foi de **24%**.

Período: Jul - Ago				
#	Lista 1		Lista 2	Suporte Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG	610 29%
	BANANA CATURA KG	=>	BATATA MONALIZA KG	610 22%
2	TOMATE L.VIDA KG	=>	BATATA MONALIZA KG	459 31%
	BATATA MONALIZA KG	=>	TOMATE L.VIDA KG	459 22%
3	ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG	432 31%
	BATATA MONALIZA KG	=>	ERVA VERDELAND 1 KG	432 20%
4	TOMATE L.VIDA KG	=>	BANANA CATURA KG	418 28%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG	418 15%
5	SAL MOC 1 KG	=>	BATATA MONALIZA KG	412 37%
	BATATA MONALIZA KG	=>	SAL MOC 1 KG	412 19%

Tabela 6: Associações no período bimestral de Julho-Agosto de 2009 com suporte relevantes
Fonte: Pesquisa.

No período bimestral de Novembro até Dezembro de 2009. De acordo com a **Tabela 8**, item 1, nota-se que a maior associação foi entre batata monaliza kg e banana catura kg. A compra associada dos produtos batata monaliza kg e banana catura kg na mesma transação somaram 613 vezes (suporte) no período. Contudo, no cálculo de divisão do suporte 613 vezes, pelos dias (25) trabalhados no mês pelo supermercado, esta compra ocorreu em média 12,26 vezes ao dia. A confiança da associação batata monaliza e banana catura foi 27%.

Descobriu-se na avaliação da regra inversa que a confiança da batata monaliza kg e banana catura kg foi de 20%. Sendo que o suporte da compra

associada dos produtos batata monaliza kg e banana catura kg possui o mesmo suporte 613 vezes.

<u>Período:</u> Set - Out						
#	Lista 1		Lista 2		Suporte	Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG		728	30%
	BANANA CATURA KG	=>	BATATA MONALIZA KG		728	24%
2	TOMATE L.VIDA KG	=>	BATATA MONALIZA KG		559	33%
	BATATA MONALIZA KG	=>	TOMATE L.VIDA KG		559	23%
3	TOMATE L.VIDA KG	=>	BANANA CATURA KG		503	30%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG		503	17%
4	ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG		496	34%
	BATATA MONALIZA KG	=>	ERVA VERDELAND 1 KG		496	20%
5	ERVA VERDELAND 1 KG	=>	BANANA CATURA KG		463	32%
	BANANA CATURA KG	=>	ERVA VERDELAND 1 KG		463	15%

Tabela 7: Associações no período bimestral de Setembro-Outubro de 2009 com suporte relevantes
Fonte: Pesquisa.

Período: Nov - Dez					
#	Lista 1		Lista 2	Suporte	Confiança
1	BATATA MONALIZA KG	=>	BANANA CATURA KG	613	27%
	BANANA CATURA KG	=>	BATATA MONALIZA KG	613	20%
2	MELANCIA ESPECIAL KG	=>	BANANA CATURA KG	548	26%
	BANANA CATURA KG	=>	MELANCIA ESPECIAL KG	548	18%
3	LARANJA KG	=>	BANANA CATURA KG	498	27%
	BANANA CATURA KG	=>	LARANJA KG	498	16%
4	ALFACE CRESPI UN	=>	BANANA CATURA KG	495	25%
	BANANA CATURA KG	=>	ALFACE CRESPI UN	495	16%
5	TOMATE L.VIDA KG	=>	BANANA CATURA KG	465	27%
	BANANA CATURA KG	=>	TOMATE L.VIDA KG	465	15%

Tabela 8: Associações no período bimestral de Novembro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

Fica demonstrado nas tabelas de análise bimestral (3 a 8) que através das regras de associação as informações descobertas tem uma correlação com a tabela anual, sugerido que o consumidor tem comportamento parecido na aquisição dos produtos associados durante o ano, confirmando a importância do setor de hortifrutigranjeiro no posicionamento dentro do supermercado, apontando produtos com confiança de até 37% (sal moc 1kg \Rightarrow batata monaliza kg). Assim, a área de marketing pode trabalhar ofertas com o perfil mais orientado ao consumo associado, facilitando a vida do consumidor e ajudando a fidelizar clientes.

4.3. Segmentos de Produtos no Supermercado - Anual

Discutir-se-á agora o período anual de Janeiro até Dezembro de 2009, tendo o foco em **grupos isolados de produtos**. Os grupos isolados são definidos como segmentos de produtos do mercado. Busca-se verificar quais são as maiores associações de compras. Justifica-se a escolha de grupo isolados pela quantidade de combinações possíveis dentro do mesmo grupo, podendo sugerir uma leitura mais adequada das análises já realizadas.

Detalhes são verificados respectivamente nas tabelas de 9 a 18. Os grupos isolados definidos pelo pesquisador são:

- Carne
- Material de Limpeza
- Refrigerante
- Açúcar
- Óleo
- Sal
- Farinha de Trigo
- Tomate.

O ponto de corte (critério) foi estabelecido através das compras com suporte acima de **500** e a confiança presente de 0 até 100%. Assim, busca-se as maiores ocorrências de frequência de compra em comportamento do consumidor, evidenciando compras associadas que foram identificadas.

Produto: **Carne - Período Jan - Dez 2009**

#	Lista 1	Lista 2	Suporte	Confiança
	PERNIL SUINO	=> BANANA CATURA KG	770	31%
	CARNE FILE AGULHA	=> BANANA CATURA KG	749	31%
	CARNE MOIDA 2 ^a	=> CEBOLA NAC. KG	739	19%
	CARNE MOIDA 2 ^a	=> ALFACE CRESPA UN	723	18%
	CARNE FILE AGULHA	=> BATATA MONALIZA KG	716	30%
	COSTELA BOVINA	=> BANANA CATURA KG	715	26%
	CARNE MOIDA 2 ^a	=> REPOLHO VERDE KG	692	18%
	CARNE MOIDA 2 ^a	=> ACUCAR CRISTAL 5 K	677	17%
	CARNE MOIDA 2 ^a	=> OVOS GALINHA CARMINA	672	17%
	PERNIL SUINO	=> TOMATE L.VIDA KG	633	25%
	COSTELA BOVINA	=> TOMATE L.VIDA KG	632	23%
	CARNE MOIDA 2 ^a	=> ESP ACO BOMBRIIL 8G	613	16%
	FILE SIMPLES	=> BANANA CATURA KG	589	30%
	FILE SIMPLES	=> BATATA MONALIZA KG	587	30%
	CARNE MOIDA 2 ^a	=> OLEO COCAMAR 900 ML	586	15%
	CARNE FILE AGULHA	=> TOMATE L.VIDA KG	577	24%
	CARNE MOIDA 2 ^a	=> LARANJA KG	573	14%
	COSTELA BOVINA	=> ALFACE CRESPA UN	556	20%
	CARNE MOIDA 2 ^a	=> COXA S/COXA GRANEL	552	14%
	PERNIL SUINO	=> ERVA VERDELAND 1 KG	526	21%

Tabela 9: Produto Carne - no período de Janeiro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

A **Tabela 9** demonstra através de suportes relevantes que o produto Carne, não importando o tipo, gera associações que podem refletir hábitos de consumo do consumidor ou ainda a forma como o supermercado direciona suas ofertas. Assim, é necessário que se interprete as informações descobertas e estas possam remeter a ações de interesse do supermercado, combinadas e direcionadas aos interesses dos consumidores.

O produto pernil suíno associado à banana catura kg obteve suporte de 770 e confiança de 31%, sendo os dois maiores índices da tabela em estudo, sugerindo que existe a probabilidade de aquisição do produto banana catura kg em 31% das operações com pernil suíno.

O produto carne moída de 2ª teve o terceiro maior suporte 739, entretanto, foi o item que associado a outros itens teve o maior destaque, aparecendo em 9 das 20 análises descobertas. Com confiança que varia de 14% a 19% entre produtos do setor hortifrutigranjeiro, açúcar, bombom, e carne.

Alguns itens, tais como: pernil suíno, carne filé agulha e filé simples, geraram uma confiança na casa dos 30% com os produtos banana caturra kg e batata monaliza kg.

O produto carne sugere que mesmo utilizando apenas a regra de $A \Rightarrow B$, descobre-se informações relevantes sobre associação de produtos. Nota-se, que o produto carne esteve associado em 15 dos 20 itens analisados com produtos do setor de hortifrutigranjeiros; carne moída 2ª é o item que mais vende dentro deste grupo; e pernil suíno gera o melhor suporte e confiança.

De acordo com a **Tabela 10**, Material de Limpeza, fica descoberto apenas dois produtos de limpeza associados e com suportes relevantes sendo, omo multi ação1 e água sanit q-bom. O Produto omo multi ação1 ficou relacionado com batata monaliza kg com suporte de **1.312** e confiança de **35%**. Na sequência ficou associado com banana caturra kg, suporte 1.186 e confiança de 32%.

Os dois produtos encontrados na sequência foram água sanit q-bom, tendo como itens associados batata monaliza kg e banana caturra kg, com suportes de 1.096 e 1.050. Ainda, confiança de 33% e 32% respectivamente.

Este achado ajuda a confirmar a importância dos produtos batata monaliza kg e banana caturra kg dentro do mix de produtos do supermercado, fato este já comentado anteriormente. Aponta que os produtos de limpeza omo multi ação1 e água sanit q-bom fazem associações com percentual de confiança relevante em termos de vendas, associando-se a uma quantidade diversificada de produtos.

Considerando as informações garimpadas na **Tabela 11** pode-se observar que o produto coca cola 2lts, 2.5lts e 1lt, teve destaque nas regras de associações encontradas, sendo a única marca de refrigerante presente nesta tabela, sugerindo a forte associação da marca a outros produtos.

Assim, associado aos produtos banana caturra kg e batata monaliza kg com os principais suportes 1.704 e 1.694 respectivamente. A confiança nos dois itens coca pet 2lt \Rightarrow banana caturra kg e coca pet 2lt \Rightarrow batata monaliza kg, foram de 23%.

O item que mais chamou atenção foi a coca cola 1lt devido a alta confiança encontrada com o produto salgados diversos, **42%**. O suporte neste produto foi de **904**. A relação encontrada suporte/confiança neste item, indica que a cada coca cola lt comercializada, existe 42% de possibilidade de um produto salgado diversos ser vendido associado. Esta informação é importante para o supermercado e deve ser analisada de forma a entender melhor o item salgados diversos, descobrindo quais salgados fazem parte deste produto, ou se é apenas um produto, e assim, conseguir estimular ainda mais a venda dos itens.

Produto: Material de Limpeza - Período Jan - Dez 2009				
#	Lista 1	Lista 2	Suporte	Confiança
	OMO MULTI ACAO 1	=> BATATA MONALIZA KG	1312	35%
	OMO MULTI ACAO 1	=> BANANA CATURA KG	1186	32%
	AGUA SANIT Q-BOA	=> BATATA MONALIZA KG	1096	33%
	AGUA SANIT Q-BOA	=> BANANA CATURA KG	1050	32%
	OMO MULTI ACAO 1	=> TOMATE L.VIDA KG	902	24%
	OMO MULTI ACAO 1	=> SAL MOC 1 KG	884	24%
	OMO MULTI ACAO 1	=> ERVA VERDELAND 1 KG	881	23%
	AGUA SANIT Q-BOA	=> ERVA VERDELAND 1 KG	771	23%
	AGUA SANIT Q-BOA	=> TOMATE L.VIDA KG	761	23%
	AGUA SANIT Q-BOA	=> SAL MOC 1 KG	744	22%
	OMO MULTI ACAO 1	=> CEBOLA NAC. KG	729	19%
	OMO MULTI ACAO 1	=> ACUCAR CRISTAL 5 K	686	18%
	OMO MULTI ACAO 1	=> REPOLHO VERDE KG	665	18%
	OMO MULTI ACAO 1	=> ESP ACO BOMBRIL 8G	655	17%
	OMO MULTI ACAO 1	=> ALFACE CRESPA UN	644	17%
	AGUA SANIT Q-BOA	=> ACUCAR CRISTAL 5 K	634	19%
	AGUA SANIT Q-BOA	=> ALFACE CRESPA UN	607	18%
	AGUA SANIT Q-BOA	=> CEBOLA NAC. KG	602	18%
	OMO MULTI ACAO 1	=> COCA PET 2 L	585	15%
	OMO MULTI ACAO 1	=> LARANJA KG	572	15%

Tabela 10: Produto Material de Limpeza - no período de Janeiro-Dezembro de 2009 com suporte relevantes

Fonte: Pesquisa.

Produto: Refrigerante - Período Jan - Dez 2009

#	Lista 1	Lista 2	Suporte	Confiança
	COCA PET 2 L	=> BANANA CATURA KG	1704	23%
	COCA PET 2 L	=> BATATA MONALIZA KG	1694	23%
	COCA PET 2,5 L	=> BANANA CATURA KG	1658	25%
	COCA PET 2,5 L	=> BATATA MONALIZA KG	1612	25%
	COCA PET 2,5 L	=> TOMATE L.VIDA KG	1276	19%
	COCA PET 2 L	=> TOMATE L.VIDA KG	1245	17%
	COCA PET 2 L	=> ALFACE CRESPA UN	1216	17%
	COCA PET 2,5 L	=> ALFACE CRESPA UN	1087	16%
	FANTA PET 2 L	=> BATATA MONALIZA KG	1068	29%
	FANTA PET 2 L	=> BANANA CATURA KG	1007	27%
	COCA PET 2 L	=> ERVA VERDELAND 1 KG	989	13%
	COCA PET 2 L	=> SAL MOC 1 KG	933	13%
	COCA PET 2,5 L	=> ERVA VERDELAND 1 KG	918	14%
	REFRIG LITRO COCA	=> SALGADOS DIVERSOS	904	42%
	COCA PET 2 L	=> CEBOLA NAC. KG	898	12%
	COCA PET 2,5 L	=> SAL MOC 1 KG	878	13%
	COCA PET 2 L	=> LARANJA KG	867	12%
	COCA PET 2,5 L	=> CEBOLA NAC. KG	844	13%
	COCA PET 2,5 L	=> REPOLHO VERDE KG	812	12%
	COCA PET 2 L	=> REPOLHO VERDE KG	796	11%

Tabela 11: Produto Refrigerante - no período de Janeiro-Dezembro de 2009 com suporte relevantes

Fonte: Pesquisa.

Examinando a **Tabela 12**, produto Açúcar, percebe-se que é uma quadro com forte suporte e confiança, talvez o argumento seja por compor a cesta básica.

Em dois momentos distintos as associações com o item açúcar se destacaram: a primeira foi associação açúcar cristal 5k \Rightarrow batata monaliza kg, dado a importância de seu suporte 1.908 e com confiança de 36%; no segundo momento açúcar cristal 2k \Rightarrow banana catura kg com suporte 1.734 e confiança 33%.

Registra-se que na tabela 12 praticamente a única marca do produto açúcar que esteve presente foi Cristal. Podendo indicar a sua força de vendas, preços competitivos, preferência do consumidor e outros. Percebe-se ainda sua forte capacidade de associação com diversos outros itens.

Produto: Açúcar - Período Jan - Dez 2009

#	Lista 1	Lista 2	Suporte	Confiança
	ACUCAR CRISTAL 5 K	=> BATATA MONALIZA KG	1908	36%
	ACUCAR CRISTAL 5 K	=> BANANA CATURA KG	1734	33%
	ACUCAR CRISTAL 5 K	=> SAL MOC 1 KG	1617	30%
	ACUCAR CRISTAL 5 K	=> ERVA VERDELAND 1 KG	1458	27%
	ACUCAR CRISTAL 5 K	=> TOMATE L.VIDA KG	1272	24%
	ACUCAR CRISTAL 5 K	=> CEBOLA NAC. KG	1193	22%
	ACUCAR CRISTAL 5 K	=> ESP ACO BOMBRIL 8G	1103	21%
	ACUCAR CRISTAL 5 K	=> REPOLHO VERDE KG	1077	20%
	ACUCAR CRISTAL 2 K	=> BANANA CATURA KG	1052	31%
	ACUCAR CRISTAL 2 K	=> BATATA MONALIZA KG	1050	31%
	ACUCAR CRISTAL 5 K	=> FERMENTO INSTANT 125	1007	19%
	ACUCAR CRISTAL 5 K	=> OLEO COCAMAR 900 ML	848	16%
	ACUCAR CRISTAL 5 K	=> ALFACE CRESPA UN	792	15%
	ACUCAR CRISTAL 5 K	=> OVOS GALINHA CARMINA	773	14%
	ACUCAR CRISTAL 2 K	=> TOMATE L.VIDA KG	718	21%
	ACUCAR CRISTAL 5 K	=> LARANJA KG	712	13%
	ACUCAR A- ALEGRE 1 K	=> BATATA MONALIZA KG	709	30%
	ACUCAR CRISTAL 5 K	=> BANANA CATURA KG, BATATA MONALIZA KG	706	13%
	ACUCAR CRISTAL 5 K	=> TRIGO KOENE 5KG	695	13%
	ACUCAR CRISTAL 5 K	=> BATATA MONALIZA KG, SAL MOC 1 KG	695	13%

Tabela 12: Produto Açúcar - no período de Janeiro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

O produto Óleo está representando pela **Tabela 13**, onde nota-se, que as associações encontradas com confiança de 30% ou mais estão atreladas a comercialização simultânea de batata monaliza kg e ou banana catura kg, corroborando com o que foi demonstrado até o momento.

Itens que são comercializados de forma associada com óleo: erva verdeland 1kg, tomate l.vida kg, sal moc 1kg, cebola nac. kg, alface crespa un, açúcar cristal 5k, repolho verde kg, esp aço Bombril 8g e laranja kg.

As marcas de produto que mais se destacaram foram: óleo cocamar 900ml, óleo soya 900ml pe e óleo coamo 900ml. A associação que mais se destacou foi

óleo Cocamar 900ml e batata monaliza kg com suporte de **1.606** e confiança de **32%**.

Produto: Óleo - Período Jan - Dez 2009				
#	Lista 1	Lista 2	Suporte	Confiança
	OLEO COCAMAR 900 ML	=> BATATA MONALIZA KG	1606	32%
	OLEO COCAMAR 900 ML	=> BANANA CATURA KG	1470	30%
	OLEO SOYA 900 ML PE	=> BATATA MONALIZA KG	1223	33%
	OLEO SOYA 900 ML PE	=> BANANA CATURA KG	1163	31%
	OLEO COCAMAR 900 ML	=> ERVA VERDELAND 1 KG	1058	21%
	OLEO COAMO 900 ML	=> BATATA MONALIZA KG	1052	33%
	OLEO COCAMAR 900 ML	=> TOMATE L.VIDA KG	1029	21%
	OLEO COCAMAR 900 ML	=> SAL MOC 1 KG	1029	21%
	OLEO COAMO 900 ML	=> BANANA CATURA KG	973	30%
	OLEO SOYA 900 ML PE	=> TOMATE L.VIDA KG	898	24%
	OLEO COCAMAR 900 ML	=> CEBOLA NAC. KG	890	18%
	OLEO COCAMAR 900 ML	=> ALFACE CRESPA UN	848	17%
	OLEO COCAMAR 900 ML	=> ACUCAR CRISTAL 5 K	848	17%
	OLEO SOYA 900 ML PE	=> SAL MOC 1 KG	821	22%
	OLEO SOYA 900 ML PE	=> ERVA VERDELAND 1 KG	818	22%
	OLEO COCAMAR 900 ML	=> REPOLHO VERDE KG	803	16%
	OLEO COAMO 900 ML	=> TOMATE L.VIDA KG	781	24%
	OLEO COCAMAR 900 ML	=> ESP ACO BOMBRIL 8G	772	15%
	OLEO COAMO 900 ML	=> SAL MOC 1 KG	753	23%
	OLEO COCAMAR 900 ML	=> LARANJA KG	711	14%

Tabela 13: Produto Óleo - no período de Janeiro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

A **Tabela 14**, composta pela análise de 20 itens do produto Sal, demonstra que 9 itens estão associadas ao setor hortifrutigranjeiro em suas vendas, 12 itens não pertencem a cesta básica, 7 itens são pertencentes a cesta básica e outros como o gás bujao p13 aparecendo pela primeira vez em uma associação..

A principal associação garimpada foi entre sal moc 1kg e batata monaliza kg, atingindo suporte de **2.362** e confiança de 35%. Em suma,, de todas as vezes que ocorreram vendas com sal moc 1kg, 35% das vendas ocorreram em conjunto batata

monaliza kg. O menor suporte encontrado foi 840, existente entre sal moc 1kg e gás bujao p13, tendo como confiança 12%.

Fica evidenciado na análise que as associações encontradas são as mais diversas possíveis. Possibilitando aos gestores uma variada e rica quantidade de combinações, em uma eventual estratégia de marketing para o produto sal.

Produto: Sal - Período Jan - Dez 2009				
#	Lista 1	Lista 2	Suporte	Confiança
	SAL MOC 1 KG	=> BATATA MONALIZA KG	2362	35%
	SAL MOC 1 KG	=> BANANA CATURA KG	2100	31%
	SAL MOC 1 KG	=> ERVA VERDELAND 1 KG	1669	25%
	SAL MOC 1 KG	=> TOMATE L.VIDA KG	1663	25%
	SAL MOC 1 KG	=> ACUCAR CRISTAL 5 K	1617	24%
	SAL MOC 1 KG	=> CEBOLA NAC. KG	1485	22%
	SAL MOC 1 KG	=> REPOLHO VERDE KG	1357	20%
	SAL MOC 1 KG	=> ESP ACO BOMBRIL 8G	1291	19%
	SAL MOC 1 KG	=> ALFACE CRESPA UN	1127	17%
	SAL MOC 1 KG	=> OLEO COCAMAR 900 ML	1029	15%
	SAL MOC 1 KG	=> FERMENTO INSTANT 125	1023	15%
	SAL MOC 1 KG	=> OVOS GALINHA CARMINA	979	14%
	SAL MOC 1 KG	=> COCA PET 2 L	933	14%
	SAL MOC 1 KG	=> LARANJA KG	917	13%
	SAL MOC 1 KG	=> TRIGO KOENE 5KG	894	13%
	SAL MOC 1 KG	=> OMO MULTI ACAA 1	884	13%
	SAL MOC 1 KG	=> BANANA CATURA KG, BATATA MONALIZA KG	883	13%
	SAL MOC 1 KG	=> COCA PET 2,5 L	878	13%
	SAL MOC 1 KG	=> COXA S/COXA GRANEL	867	13%
	SAL MOC 1 KG	=> GAS BUJAO P 13	840	12%

Tabela 14: Produto Sal - no período de Janeiro-Dezembro de 2009 com suporte relevantes

Fonte: Pesquisa.

Produto: Farinha de Trigo - Período Jan - Dez 2009				
#	Lista 1	Lista 2	Suporte	Confiança
	TRIGO KOENE 5KG	=> BATATA MONALIZA KG	1109	37%
	TRIGO KOENE 5KG	=> BANANA CATURA KG	978	33%
	TRIGO ANACONDA 5 KG	=> BATATA MONALIZA KG	966	37%
	TRIGO KOENE 5KG	=> SAL MOC 1 KG	894	30%
	TRIGO ANACONDA 5 KG	=> BANANA CATURA KG	890	34%
	TRIGO KOENE 5KG	=> ERVA VERDELAND 1 KG	809	27%
	TRIGO ANACONDA 5 KG	=> SAL MOC 1 KG	774	29%
	TRIGO ANACONDA 5 KG	=> ERVA VERDELAND 1 KG	765	29%
	TRIGO ANACONDA 1KG	=> BATATA MONALIZA KG	746	32%
	TRIGO KOENE 5KG	=> TOMATE L.VIDA KG	737	25%
	TRIGO ANACONDA 1KG	=> BANANA CATURA KG	715	30%
	TRIGO KOENE 5KG	=> CEBOLA NAC. KG	696	23%
	TRIGO KOENE 5KG	=> ACUCAR CRISTAL 5 K	695	23%
	TRIGO ANACONDA 5 KG	=> ACUCAR CRISTAL 5 K	667	25%
	TRIGO KOENE 5KG	=> REPOLHO VERDE KG	637	21%
	TRIGO ANACONDA 5 KG	=> TOMATE L.VIDA KG	615	23%
	TRIGO KOENE 5KG	=> FERMENTO INSTANT 125	608	20%
	TRIGO KOENE 5KG	=> ESP ACO BOMBRIL 8G	598	20%
	TRIGO ANACONDA 5 KG	=> ESP ACO BOMBRIL 8G	572	21%
	TRIGO ANACONDA 5 KG	=> CEBOLA NAC. KG	564	21%

Tabela 15: Produto Farinha de Trigo - no período de Janeiro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

Realizar-se-á agora a análise da **Tabela 15**, sendo o ponto chave o produto: Farinha de Trigo. Neste estudo o destaque das associações ficaram com duas marcas de trigo, Koene e Anaconda. Independente do tamanho da embalagem, 5kg ou 1kg, quando associadas, a venda com os produtos batata monaliza kg e banana caturra kg vai de **30% a 37%**. Potencializando-se a importância destes produtos no portfólio de produtos do supermercado e fortalecendo a sua força de vendas associadas a outros tipos de itens tais como: sal moc 1kg, erva verdeland 1kg, tomate l.vida kg, cebola nac.kg, açúcar cristal 5k, repolho verde kg, fermento instant 125 e esp aço bombril 8g.

Ficou salientado que a associação trigo koene 5kg com batata monaliza kg, gerou suporte de 1.109 e confiança de 37%. De igual maneira aparece a venda associada trigo anaconda 5kg com batata monaliza kg, gerando confiança de 37%, apenas com suporte um pouco menor 978, identificando que as duas marcas possuem confiança significativa.

Neste grupo examinado fica evidente que a variável confiança é forte em todas as associações descobertas, sendo que a menor confiança encontrada ficou em 21%. Este fato pode estar atrelado a importância do produto farinha de trigo no dia a dia do consumidor, através de vários produtos que podem ser desenvolvidos, principalmente para consumo.

<u>Produto:</u> Tomate - Período Jan - Dez 2009				
#	Lista 1	Lista 2	Suporte	Confiança
	TOMATE L.VIDA KG	=> BATATA MONALIZA KG	2921	29%
	TOMATE L.VIDA KG	=> BANANA CATURA KG	2866	28%
	TOMATE L.VIDA KG	=> ALFACE CRESPA UN	2109	20%
	TOMATE L.VIDA KG	=> CEBOLA NAC. KG	1735	17%
	TOMATE L.VIDA KG	=> ERVA VERDELAND 1 KG	1721	17%
	TOMATE L.VIDA KG	=> REPOLHO VERDE KG	1695	16%
	TOMATE L.VIDA KG	=> SAL MOC 1 KG	1663	16%
	TOMATE L.VIDA KG	=> LARANJA KG	1477	14%
	TOMATE L.VIDA KG	=> OVOS GALINHA CARMINA	1339	13%
	TOMATE L.VIDA KG	=> COCA PET 2,5 L	1276	12%
	TOMATE L.VIDA KG	=> ACUCAR CRISTAL 5 K	1272	12%
	TOMATE L.VIDA KG	=> COCA PET 2 L	1245	12%
	TOMATE L.VIDA KG	=> ESP ACO BOMBRIIL 8G	1208	12%
	TOMATE L.VIDA KG	=> VERD. PC DIVERSA	1169	11%
	TOMATE L.VIDA KG	=> MAMAO FORMOSA	1162	11%
	TOMATE L.VIDA KG	=> MELANCIA ESPECIAL KG	1052	10%
	TOMATE L.VIDA KG	=> OLEO COCAMAR 900 ML	1029	10%
	TOMATE L.VIDA KG	=> BANANA CATURA KG, BATATA MONALIZA KG	1016	10%
	TOMATE L.VIDA KG	=> COXA S/COXA GRANEL	995	9%
	TOMATE L.VIDA KG	=> CARNE MOIDA 2 ^a	977	9%

Tabela 16: Produto Tomate - no período de Janeiro-Dezembro de 2009 com suporte relevantes
Fonte: Pesquisa.

A **tabela 16** demonstra a importância do produto tomate l.vida kg no estudo do *Data Mining*, pois três informações que se destacam:

A primeira acerca da variedade de itens que se correlacionam, sendo: batata monaliza kg, banana caturra kg, erva verdeland 1kg, sal moc 1kg, cebola nac. kg, alface crespa un, açúcar cristal 5k, repolho verde kg, esp aço Bombril 8g, ovos galinha carmina, coca pet 2lt e 2,5 lt, laranja kg, verd. pc diversa, mamão formosa, óleo Cocamar 900ml, coxa s/coxa granel, carne moída 2ª, melancia especial;

A segunda informação relevante é o suporte, dada a quantidade de vezes que ocorre as associações em todas as análises, iniciando em 977 e tendo como maior suporte 2.921; e

A terceira se refere a confiança que os produtos batata monaliza kg e banana caturra kg surgem, 29% e 28% respectivamente.

4.3. Análise tripla: 3 produtos – Anual

A seguir é descrito os resultados das tabelas 17 e 18, que ficaram associadas com três itens. Sendo que a tabela 17 será descrita pela variável suporte e a tabela 18 pela variável confiança.

Associada à mais dois itens, a **Tabela 17**, utiliza-se da regra $A, B \Rightarrow C$ ainda, $A, C \Rightarrow B$ e $C, B \Rightarrow A$. Assim, descobriu-se com 1.016 (suporte) os itens associados, banana catura kg + batata monaliza kg \Rightarrow tomate l.vida kg e com confiança de 26%, invertendo a associação ficou banana catura kg + tomate l.vida kg \Rightarrow batata monaliza kg obtendo confiança de 35% e por terceiro a associação tomate l.vida kg + batata monaliza kg \Rightarrow banana catura kg 34% de confiança.

Percebe-se nesta análise que o **trio** (batata+banana+tomate) que são itens associados pelo consumidor a todo momento e que juntos somam percentual significativo na representação do conjunto de produtos associados da empresa pesquisada. Na busca de conhecimento do comportamento de compras do consumidor poderia-se pesquisar com mais detalhes a combinação deste trio de produtos, com o objetivo de incentivar o consumo associado.

Produto: Diversos - Período Jan - Dez 2009

#	Lista 1	Lista 2	Suporte	Confiança
	BANANA CATURA KG, BATATA MONALIZA KG	=> TOMATE L.VIDA KG	1016	26%
	BANANA CATURA KG, TOMATE L.VIDA KG	=> BATATA MONALIZA KG	1016	35%
	TOMATE L.VIDA KG, BATATA MONALIZA KG	=> BANANA CATURA KG	1016	34%
	BANANA CATURA KG, BATATA MONALIZA KG	=> ERVA VERDELAND 1 KG	942	24%
	BANANA CATURA KG, ERVA VERDELAND 1 KG	=> BATATA MONALIZA KG	942	39%
	BATATA MONALIZA KG, ERVA VERDELAND 1 KG	=> BANANA CATURA KG	942	36%
	BANANA CATURA KG, BATATA MONALIZA KG	=> SAL MOC 1 KG	883	23%
	BANANA CATURA KG, SAL MOC 1 KG	=> BATATA MONALIZA KG	883	42%
	BATATA MONALIZA KG, SAL MOC 1 KG	=> BANANA CATURA KG	883	37%
	ALFACE CRESPI UN, BATATA MONALIZA KG	=> BANANA CATURA KG	852	34%
	BANANA CATURA KG, ALFACE CRESPI UN	=> BATATA MONALIZA KG	852	30%
	BANANA CATURA KG, BATATA MONALIZA KG	=> ALFACE CRESPI UN	852	22%
	BANANA CATURA KG, BATATA MONALIZA KG	=> REPOLHO VERDE KG	819	21%
	BANANA CATURA KG, REPOLHO VERDE KG	=> BATATA MONALIZA KG	819	38%
	REPOLHO VERDE KG, BATATA MONALIZA KG	=> BANANA CATURA KG	819	38%
	BANANA CATURA KG, BATATA MONALIZA KG	=> CEBOLA NAC. KG	794	20%
	BANANA CATURA KG, CEBOLA NAC. KG	=> BATATA MONALIZA KG	794	38%
	BATATA MONALIZA KG, CEBOLA NAC. KG	=> BANANA CATURA KG	794	36%

Tabela 17: Produto Diversos - no período de Janeiro-Dezembro de 2009 com três itens associados e com suporte relevantes

Fonte: Pesquisa.

Aplicando apenas a regra $A, B \Rightarrow C$, a **Tabela 18** foi analisada pela variável confiança. Nota-se, que as três primeiras associações mostraram a confiança de 45%, indicando que a cada cem transações com os itens associados, existe a possibilidade de que o consumidor adquira em 45% das vezes batata monaliza kg. Percebe-se que a menor confiança encontrada nesta tabela ficou em 42%, sugerindo que associações com mais de dois itens são viáveis e com vasta possibilidade de combinações.

A variável suporte se posicionou, ver tabela 18, entre 529 e 883, sinalizando que um trabalho de prospecção pode ser desenvolvido para alavancar esses números, uma vez que a confiança descoberta é relevante.

Produto: Diversos - Período Jan - Dez 2009				
#	Lista 1		Lista 2	Suporte Confiança
	ESP ACO BOMBRIL 8G, SAL MOC 1 KG	=>	BATATA MONALIZA KG	589 45%
	ESP ACO BOMBRIL 8G, ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG	577 45%
	TOMATE L.VIDA KG, ACUCAR CRISTAL 5 K	=>	BATATA MONALIZA KG	576 45%
	SAL MOC 1 KG, ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG	750 44%
	ACUCAR CRISTAL 5 K, CEBOLA NAC. KG	=>	BATATA MONALIZA KG	534 44%
	TOMATE L.VIDA KG, SAL MOC 1 KG	=>	BATATA MONALIZA KG	726 43%
	SAL MOC 1 KG, CEBOLA NAC. KG	=>	BATATA MONALIZA KG	653 43%
	ERVA VERDELAND 1 KG, CEBOLA NAC. KG	=>	BATATA MONALIZA KG	641 43%
	REPOLHO VERDE KG, SAL MOC 1 KG	=>	BATATA MONALIZA KG	587 43%
	BANANA CATURA KG, SAL MOC 1 KG	=>	BATATA MONALIZA KG	883 42%
	TOMATE L.VIDA KG, ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG	734 42%
	SAL MOC 1 KG, ACUCAR CRISTAL 5 K	=>	BATATA MONALIZA KG	695 42%
	ERVA VERDELAND 1 KG, ACUCAR CRISTAL 5 K	=>	BATATA MONALIZA KG	620 42%
	REPOLHO VERDE KG, ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG	590 42%
	ALFACE CRESPA UN, ERVA VERDELAND 1 KG	=>	BATATA MONALIZA KG	535 42%
	BANANA CATURA KG, ESP ACO BOMBRIL 8G	=>	BATATA MONALIZA KG	672 41%
	MAMAO FORMOSA, BATATA MONALIZA KG	=>	BANANA CATURA KG	614 41%
	BANANA CATURA KG, LEITE COND FRIMESA	=>	BATATA MONALIZA KG	529 41%

Tabela 18: Produto Diversos - no período de Janeiro-Dezembro de 2009 com três itens associados e com confiança relevantes

Fonte: Pesquisa

4.4 Combos Sugeridos ao Varejista

Foi sugerido ao supermercado Franzoni, conforme objetivo específico, expor através de “pontos de gôndola”, **quatro combos de vendas**.

Os combos foram formados por dois itens associados, sendo que o **pesquisador** solicitou a inclusão de dois produtos: batata monaliza kg e banana caturra kg, devido aos números descobertos em análises já comentadas. Os demais produtos foram escolhas de **interesse do supermercado** e sem critérios pré-

definidos. O acompanhamento pelo pesquisador foi semanal durante o período de exposição dos combos.

O período de exposição dos combos foi de trinta (30) dias, durante o mês de Maio e Junho de 2010 para que se entenda o comportamento mensal do consumidor no supermercado, sem levar em conta o período (sazonalidade) do mês pesquisado ou ainda períodos festivos como páscoa, dia das mães, dias dos pais e outros.

Busca-se através desta estratégia de marketing uma melhor interpretação dos dados até aqui descritos, e maior conhecimento do comportamento de compras associadas do consumidor, criando a possibilidade de comparação das associações antes e depois dos combos. Tudo, com objetivo de gerar conhecimentos que possam dar suporte ao gestor na tomada de decisões, definir estratégias de marketing e ainda possibilitar a garimpagem de novas oportunidades de negócios. Os Combos sugeridos foram:

I) 2 kg Batata Monaliza + 1un Pêssego Metades em Calda Cantu lata 400gr

II) 2 Kg Banana Caturra + 1un doce de leite Cantu 400gr

III) 1un Detergente Lava Louças Ype Neutro 500ml + 1un Q-bou Multiuso Frutas Cítricas 500ml

IV) 1un Tagliarini Caseiro Dybom 500gr + 1un Biscoito Águia Salt Cracker 360gr

Análise dos dados dos combos. O primeiro combo proposto e analisado estava assim descrito: **Combo I** - 2 kg batata monaliza + 1un pêssego metades em calda cantu lata 400gr. De acordo com o *Data Mining*, percebe-se, na tabela 19, que o produto batata monaliza kg teve uma **redução** nas transações de aproximadamente 10% no período comparado. Enquanto o item pêssego calda 400g teve um **acréscimo** significativo, **770%** no número de transações no mesmo período.

Avaliando a associação batata \Rightarrow pêssego, temos em Maio/09 suporte igual a 10 e no período Maio/10 suporte de 77. Não teve confiança no mês de Maio/09,

chegando a 4% em Maio/10. Na regra inversa, pêssego \Rightarrow batata, o suporte permanece (10) e a confiança fica em 59% no mês de Maio/09 e 52% no mês Maio/10 o que é expressivo.

Produtos	Maio/09		Maio/10	
	Transações		Transações	
Batata Monaliza Kg	2.028		1.838	
Pessegue Calda 400 G.Metades Cantu	17		148	
Associações	Suporte	Confiança	Suporte	Confiança
Batata => Pessegue	10	0%	77	4%
Pessegue => Batata	10	59%	77	52%

Tabela 19: Análise Combo I - período de Maio/09 e Maio/10
Fonte: Pesquisa.

Na tabela 20, o período comparado é Maio-Junho/09 com Maio-Junho/10. Com relação as transações, obteve-se um aumento de 6% do produto batata monalisa kg, e o item pêssego um crescimento próximo a 5.400%.

Nas associações, no que tange a suporte e confiança, no primeiro período, batata \Rightarrow pêssego, ambos ficaram com o valor zero. No segundo período, suporte de 64 e confiança 3%.

Observa-se que na associação inversa pêssego \Rightarrow batata, permanece suporte e confiança zerados no período Maio-Junho/09. Avaliando o período Maio-Junho/10, tem-se o suporte 64 e confiança 58%.

Produtos	Maio-Junho/09		Maio-Junho/10	
	Transações		Transações	
Batata Monaliza Kg	1.764		1.875	
Pessegue Calda 400 G.Metades Cantu	2		110	
Associações	Suporte	Confiança	Suporte	Confiança
Batata => Pessegue	0	0%	64	3%
Pessegue => Batata	0	0%	64	58%

Tabela 20: Análise Combo I - período de Maio-Junho/09 e Maio-Junho/10
Fonte: Pesquisa.

No período de Maio/10, o item pesquisado batata monalisa kg sofreu interferência significativa no preço, oscilando entre R\$ 1,50 e 2,60 (fonte Dieese) o quilograma, devido as fortes chuvas que aconteceram no período, uma variação do valor de venda de 74%. Já no mês de Maio-Junho/10, o preço volta ao valor normal de mercado, refletindo na quantidade de transações descobertas. Verificando tanto suporte como confiança nos períodos de Maio/10 e Maio-Junho/10 tem-se que as associações lograram êxito convincente. A foto 7, demonstra o Combo I exposto no supermercado.



Foto 7: Combo I em exposição no supermercado

Interpretado na sequência, o **Combo II**, formado pelos itens, 2 Kg banana caturra + doce de leite Cantu 400gr, indicam que no período comparado, tabela 21, existe uma redução nas transações de 2% no produto banana caturra e um acréscimo cerca de 65% no produto doce de leite Cantu 400gr.

Com relação as associações sugeridas, a variável suporte atingiu 53% de acréscimo e a variável confiança cresceu na associação doce \Rightarrow banana o equivalente a 50%. Verificando a regra inversa, temos um incremento próximo de 41%.

Produtos	Maio/09		Maio/10	
	Transações		Transações	
Banana Catura Kg	2.586		2.527	
Doce 400 G. Leite Cantu	91		151	
Associações	Suporte	Confiança	Suporte	Confiança
Doce => Banana	47	2%	72	3%
Banana => Doce	47	52%	72	48%

Tabela 21: Análise Combo II - período de Maio/09 e Maio/10
Fonte: Pesquisa

Analisando a tabela 22, percebe-se que o número de transações oscilaram em 55 para menos no segundo período do produto banana caturra kg e um acréscimo de 70 para 100 no doce de leite Cantu 400gr.

O suporte se manteve praticamente igual nos dois períodos. Já a confiança na regra doce => banana, tanto em Maio-Junho/09, como, Maio-Junho/10, foi igual a 2%. Invertendo a regra para banana => doce, houve uma redução percentual de 17% na confiança.

Produtos	Maio-Junho/09		Maio-Junho/10	
	Transações		Transações	
Banana Catura Kg	2.379		2.324	
Doce 400 G. Leite Cantu	70		100	
Associações	Suporte	Confiança	Suporte	Confiança
Doce => Banana	37	2%	36	2%
Banana => Doce	37	53%	36	36%

Tabela 22: Análise Combo II - período de Maio-Junho/09 e Maio-Junho/10
Fonte: Pesquisa.

Sintetizando, o Combo II, demonstra um **equilíbrio** nas transações com item banana catura kg e um acréscimo elevado no item doce de leite Cantu 400gr. Fica demonstrado a importância que se obteve no crescimento das vendas com a associação estudada. O Combo II pode ser visualizado na foto 8.



Foto 8: Combo II em exposição no supermercado

Combo III, combinado pelos itens, 1 q-bon multiuso frutas cítricas 500ml + 1 detergente lava louças ype neutro 500ml, conforme descrito na tabela 23, demonstra aumentos expressivos em termos de transações, 628% no item limp multi uso q-bon e 72% no detergente lava louças ype. O suporte foi de zero em Maio/09 para 35 em Maio/10, tanto na regra $A \Rightarrow B$, como $B \Rightarrow A$.

A confiança na associação $\text{limp} \Rightarrow \text{detergente}$ que era de zero, chegou a 69%, e na regra inversa de zero foi a 17%.

Produtos	Maio/09		Maio/10	
	Transações		Transações	
Limp Multi Uso 500 Ml Fta Citr q Boa	7		51	
Deterg Liq 500 Ml Neutro Ype	121		209	
Associações	Suporte	Confiança	Suporte	Confiança
Limp => Detergente	0	0%	35	69%
Detergente => Limp	0	0%	35	17%

Tabela 23: Análise Combo III - período de Maio/09 e Maio/10
Fonte: Pesquisa.

Na tabela 24, percebe-se da mesma forma o aumento convincente no número de transações, que no item limp multi uso q-bou variou de 1 para 85 (**8.400%**) e no item detergente lava louças ype de 79 a 175 (121%).

Nas associações limp => detergente e detergente => limp, no período Maio-Junho/09, as variáveis suporte e confiança estavam zerados. No período Maio-Junho/10, o suporte para ambos subiu para 65.

A confiança na associação limp => detergente cresceu 76% e nos itens detergente => limp 37%.

Produtos	Maio-Junho/09		Maio-Junho/10	
	Transações		Transações	
Limp Multi Uso 500 Ml Fta Citr q Boa	1		85	
Deterg Liq 500 Ml Neutro Ype	79		175	
Associações	Suporte	Confiança	Suporte	Confiança
Limp => Detergente	0	0	65	76%
Detergente => Limp	0	0%	65	37%

Tabela 24: Análise Combo III - período de Maio-Junho/09 e Maio-Junho/10

Fonte: Pesquisa.

No Combo III, formados pelos itens - 1 detergente lava louças ype neutro 500ml + 1 q-bou multiuso frutas cítricas 500ml. Percebe-se claramente nos dois períodos comparativos um **aumento** de transações de ambos os produtos, tanto a nível individual quanto de forma associada. Sugerindo que as informações obtidas através da ferramenta *Data Mining*, quando colocadas em prática, podem resultar em um conhecimento maior a respeito do consumo de produtos associados, possibilitando ao gestor utilizar estratégias mais adequadas ao perfil do consumidor, atraindo e fidelizando clientes, com objetivo de impulsionar as vendas.

A foto 9 vislumbra a oferta do Combo III.



Foto 9: Combo III em exposição no supermercado

Último combo sugerido, e implantado, foi o **Combo IV** - 1 biscoito água salt cracker 360gr + 1 tagliarini caseiro dybom 500gr.

Percebe-se, ver na tabela 25, que ocorreu uma redução nas transações individuais de ambos os produtos entre os períodos comparativos, sendo que no biscoito água foi de 10% e no tagliarini caseiro dybom foi de aproximadamente 25%.

Verificando a associação biscoito \Rightarrow massa, a variável suporte ficou em 5 e a confiança foi de 5%, no período Maio/09. Já no período Maio/10, o suporte subiu para 15, ou seja, maior em 200% e a confiança foi de 5% para 14%, um acréscimo de 180% .

Na regra inversa, massa \Rightarrow biscoito, que tem igual suporte 15, contudo, a confiança sobe de 9% para 34%, tendo uma variação incremental de 277%.

Produtos	Maio/09		Maio/10	
	Transações		Transações	
Bisc Salt 360 G. Tradicional Aguaia	101		111	
Massa C/ Ovos 500 G. Cas N° 2 Dybom	55		44	
Associações	Suporte	Confiança	Suporte	Confiança
Biscoito \Rightarrow Massa	5	5%	15	14%
Massa \Rightarrow Biscoito	5	9%	15	34%

Tabela 25: Análise Combo IV - período de Maio/09 e Maio/10

Fonte: Pesquisa

No período entre Maio-Junho/09 e Maio-Junho/10, ver tabela 26, o biscoito água e tagliarini caseiro dybom tiveram comportamentos diferentes. O biscoito água teve uma redução próxima de 30% nas transações e o tagliarini caseiro dybom teve um acréscimo cerca de 91%.

A associação biscoito \Rightarrow massa teve suporte igual a 2 no primeiro período e 23 no segundo, implicando em um aumento de 1.050%. Em relação a confiança, no primeiro período foi de 2% e no segundo de 27%, resultando em aumento de 1.250%.

A inversão da regra em estudo, massa \Rightarrow biscoito, tem suporte igual a análise anterior 2 e 23 nos períodos Maio-Junho/09 e Maio-Junho/10 e na confiança de 4% no período Maio-Junho/09 e 26% no período Maio-Junho/10, sugerindo um aumento de 550%,

.Produtos	Maio-Junho/09		Maio-Junho/10	
	Transações		Transações	
Bisc Salt 360 G. Tradicional Água	110		85	
Massa C/ Ovos 500 G. Cas N° 2 Dybom	47		90	
Associações	Suporte	Confiança	Suporte	Confiança
Biscoito \Rightarrow Massa	2	2%	23	27%
Massa \Rightarrow Biscoito	2	4%	23	26%

Tabela 26: Análise Combo IV - período de Maio-Junho/09 e Maio-Junho/10

Fonte: Pesquisa.

Os resultados encontrados no Combo IV fortalecem as regras de associação, sugeridos pelo aumento das variáveis suporte e confiança. De posse destas informações, é possível que o supermercado trabalhe com uma visão mais lapidada na interpretação do consumo. Assim, percebe a importância de associar itens na estratégia de incrementar vendas.

A foto 10 demonstra o Combo IV em exposição.



Foto 10: Combo IV em exposição no supermercado

4.5 RESUMO DAS VARIAÇÕES NOS COMBOS PESQUISADOS

A tabela 27 salienta um resumo de todas as operações realizadas no período de Maio/09 e Maio/10 e suas respectivas variações, abordando quantidade de transação, percentual de participação, volume e variação. Assim, realizar-se-á uma interpretação da tabela 27 nos itens descritos a seguir.

Todos os produtos movimentaram no período de Maio/09 um total de 24.217 transações. No período de Maio/10 as transações ficaram em 24.593, praticamente sem oscilações em termos percentuais no período.

O produto **banana catura kg** teve em Maio/09 2.586 transações, correspondendo a 10,68% das operações totais. No mês de Maio/10 as operações somaram 2.527, correspondendo a 10,28% do total, ou seja, menos 2,28% em relação ao ano anterior, respectivamente 59 operações a menos.

A banana catura kg sugere um alto grau de comercialização, evidenciado pelo percentual de 10% na participação total das transações ocorridas no período pesquisado, posicionando sua importância dentro dos itens adquiridos pelos

consumidores e possibilitando a criação de novas associações para auferir o potencial deste item de alavancar novos negócios.

O item **batata monaliza kg** atingiu em Maio/09 2.028 transações, o que corresponde a 8,37% das operações totais. No período Maio/10, as transações atingiram 1.838, representando 7,47% da totalidade das operações, desta forma, menor em 9,37% em comparação com o período anterior, resultando um decréscimo de 190 operações.

Avaliando as tabelas de 2 até a 18, ainda que tenha apresentando uma variação **negativa** na quantidade de transações no período comparado, conforme tabela 27, fica evidenciado a importância do produto batata monaliza kg dentro do supermercado. Ocorre que no mês de maio de 2010, devido a situações climáticas e excesso de chuvas, o preço do produto, que era em média R\$ 1,50, passou para R\$ 2,60 afetando conseqüentemente o consumo do produto.

Produtos	Maio/09		Maio/10		Variação	
	Transações	Participação	Transações	Participação	Volume	%
Todas	24.217	100,00%	24.593	100,00%	-	-
BANANA CATURA KG	2.586	10,68%	2.527	10,28%	-59	-2,28%
BATATA MONALIZA KG	2.028	8,37%	1.838	7,47%	-190	-9,37%
Bisc Salt 360 G. Tradicional Águia	101	0,42%	111	0,45%	10	9,90%
Deterg Liq 500 Ml Neutro Ype	121	0,50%	209	0,85%	88	72,73%
Doce 400 G. Leite Cantu	91	0,38%	151	0,61%	60	65,93%
Limp Multi Uso 500 Ml Fta Citr q Boa	7	0,03%	51	0,21%	44	628,57%
Massa C/ Ovos 500 G. Cas N° 2 Dybom	55	0,23%	44	0,18%	-11	-20,00%
Pessegue Calda 400 G. Metades Cantu	17	0,07%	148	0,60%	131	770,59%

Tabela 27: Resumo de todas operações realizadas e variações - período de Maio/09 e Maio/10
Fonte: Pesquisa.

O **Bisc Salt 360g. Tradicional Águia** apurou em Maio/09 101 transações, correspondendo a 0,42% das operações totalizadas. Em Maio/10, as operações

subiram para 111, representando 0,45% das transações. Com 10 transações a mais em Maio/10, o crescimento chegou a 9,90%.

Pela análise realizada, sugere-se que o produto biscoito salt água, associado com tagliarini caseiro dybom, obteve um resultado superior, próximo a 10% do que teria sem ser colocado em ponta de gôndola, indicando que o produto correlatado, quando bem posicionado, pode fazer diferença no volume de vendas.

Com 121 transações em Maio/09 o item **Deterg Liq 500MI Neutro Ype** representou 0,50% das operações totalizadas. No período Maio/10 apurou-se 209 transações, equivalentes a 0,85% das vendas totais. Comparando os períodos, apurou-se um acréscimo de 88 transações e ou 72%.

O item deterg liq 500ml neutro ype, individualmente, tinha uma comercialização de 121 unidades, conforme indicado na tabela 27. Quando foi associado ao produto limp multi uso 500ml fta citr Q-boua este atingiu resultados ainda maiores. Esta combinação tende sugerir que, através de itens associados e destacados na área de venda, podem despertar o interesse do consumidor e por consequência aumento do consumo.

O produto **Doce 400G. Leite Cantu** atingiu em Maio/09 91 transações, o equivalente a 0,38% das operações totais. Já, em Maio/10 as operações somaram 151, correspondendo a 0,61%, representando uma evolução nas vendas de 65,93% em relação ao ano anterior.

Associado com o item banana catura kg, líder em transações, conforme demonstra tabela 27, o produto doce 400g. leite cantu teve aumento significativo de vendas. Indicando que a associação resultou positivamente.

A tabela 27, no que se refere ao produto **Limp Multi Uso 500Mi Fta Citr Q-Boua**, no período Maio/09, indicou a existência de 7 transações, equivalendo 0,03% das operações totais. As transações no período Maio/10 somaram 51, em percentual equivale a 0,21% do volume total. Comparando os períodos, tem-se um crescimento de 628%. Alavancado com a venda associada ao item deterg liq 500ml neutro ype, fica demonstrado o alcance das regras de associação.

Com participação de 0,23% no total das vendas, o produto **Massa C/ Ovos 500G Cas N.2 Dybom** aparece em Maio/09 com 55 transações. No período Maio/10 as operações com o mesmo item somaram 44 operações, o equivalente a 0,18% das vendas. Quando comparado o período Maio/09 e Maio/10 identifica-se uma retração das vendas totais na ordem de 20%.

Indica a tabela 27 uma variação negativa de 20% na venda do item massa c/ ovos 500g cas nº 2 dybom no período comparado. Entretanto, se for avaliado a tabela 25 Combo IV, percebe-se que existiu um acréscimo tanto no suporte como na confiança associado ao item bisc salt 360g. tradicional água, revelando a necessidade de interpretação adequada das informações garimpadas.

No período pesquisado Maio/09 e Maio/10 o produto **Pêssego Calda 400G. Metades Cantu** obteve 17 e 148 transações respectivamente. No primeiro período correspondeu a 0,07% do total de vendas e no segundo período respondeu por 0,60% das vendas totais, sugerindo elevação de 770% em termos de comercialização.

Nota-se que o item teve um incremento de 770% quando associado ao item bata monaliza kg, conforme indica a tabela 17 Combo I, e que se o item pêssego calda 400g. metades cantu continuasse sem nenhum incentivo, a venda poderia seguir os patamares iniciais, desperdiçando o supermercado a oportunidade de elevar vendas e por consequência rentabilidade.

A seguir será analisada a tabela 28 que promove um resumo de todas as operações realizadas no período Maio-Junho/09 e Maio-Junho/10 com suas respectivas variações, demonstrando quantidade de transações, percentual de participação, volume e variação.

Os produtos movimentaram, no período de Maio-Junho/09 24.146 transações. No período Maio-Junho/10 as transações finalizaram com um total de 23.687. Comparando os dois períodos tem-se resultados muito próximos.

Conforme tabela 28, o produto **banana catura kg** teve em Maio-Junho/09 2.379 transações, correspondendo a 9,85% das operações totais. Já, em Maio-Junho/10 as operações somaram 2.324, correspondendo a 9,81% do total, ou seja, 2,31% menor que em relação ao ano anterior.

O item banana catura kg sugere um alto grau de comercialização, evidenciado pelo percentual próximo a 10% na participação total das transações ocorridas no período pesquisado, reforçando sua importância dentro dos itens adquiridos pelos consumidores e possibilitando a criação de novas associações para elevar o potencial deste item de alavancar novos negócios.

Produtos	Maio-Junho/09		Maio-Junho/10		Variação	
	Transações	Participação	Transações	Participação	Volume	%
Todas	24.146	100,00%	23.687	100,00%	-	-
BANANA CATURA KG	2.379	9,85%	2.324	9,81%	-55	-2,31%
BATATA MONALIZA KG	1.764	7,31%	1.875	7,92%	111	6,29%
Bisc Salt 360 G. Tradicional Águia	110	0,46%	85	0,36%	-25	-22,73%
Deterg Liq 500 Ml Neutro Ype	79	0,33%	175	0,74%	96	121,52%
Doce 400 G. Leite Cantu	70	0,29%	100	0,42%	30	42,86%
Limp Multi Uso 500 Ml Fta Citr q Boa	1	0,00%	85	0,36%	84	8400,00%
Massa C/ Ovos 500 G. Cas Nº 2 Dybom	47	0,19%	90	0,38%	43	91,49%
Pêssego Calda 400 G. Metades Cantu	2	0,01%	110	0,46%	108	5400,00%

Tabela 28: Resumo de todas operações realizadas e variações – período Maio-Junho/09 e Maio-Junho/10

Fonte: Pesquisa.

O item **batata monaliza kg** atingiu em Maio-Junho/09 1.764 transações, correlato a 7,31% das operações totais. No período Maio-Junho/10 as transações atingiram 1.875, representando 7,92% do total das operações.

De acordo com a tabela 28, nota-se que o produto batata monaliza kg sofre menos a interferência do clima, resultando em variação positiva de 6,29%, destacando a importância do produto batata monaliza kg dentro do supermercado.

O **Bisc Salt 360g. Tradicional Águia** apurou em Maio-Junho/09, 110 transações, correspondendo a 0,46% das operações totalizadas. Em Maio-Junho/10 as operações subiram para 85, representando 0,36% das transações. Com 25 transações a menos em Maio-Junho/10 a queda nas vendas atingiram 22,73%.

Pela análise realizada na tabela 28, percebe-se que houve uma queda na comercialização do item bisc salt 360g. tradicional águia e que pode ser sugerida através do excesso de exposição no mesmo local e talvez por tempo demasiado. Numa próxima ação poderá ser determinado um prazo menor de exposição do produto para poder comparar melhor os resultados.

Com 79 transações em Maio-Junho/09, o item **Deterg Liq 500MI Neutro Ype**, representou 0,33% das operações totalizadas. No período Maio-Junho /10 apurou-

se 175 transações, equivalentes a 0,74% das vendas totais. Comparando os períodos, apurou-se um acréscimo de 96 transações e ou 121%.

O item deterg liq 500ml neutro ype individualmente tinha uma comercialização de 79 unidades, conforme indicado na tabela 28. Quando foi associado ao produto limp multi uso 500ml fta citr q-boua, atingiu resultados ainda maiores. Esta combinação tende sugerir que através de itens associados e destacados na área de venda podem refletir a necessidade do consumidor e por conseqüência aumento do consumo.

O produto **Doce 400G. Leite Cantu** atingiu em Maio-Junho/09 70 transações, equivalente a 0,29% das operações totais. Já, em Maio-Junho/10 as operações somaram 100, correspondendo a 0,42%, representando uma evolução nas vendas de 42,86% em relação ao ano anterior.

Associado com o item banana catura kg, líder em transações, conforme demonstra tabela 28, o produto doce 400g. leite cantu teve aumento significativo de vendas, idicando que a associação resultou em sucesso.

A tabela 28, no que se refere ao produto **Limp Multi Uso 500Mi Fta Citr Q-Boa** no período Maio-Junho/09, indicou a existência de 1 transação, equivalendo a 0% das operações totais. As transações no período Maio-Junho/10 somaram 85, em percentuais equivale a 0,36% do volume total. Comparando os períodos tem-se um crescimento de 8400%.

Alavancado com a venda associada ao item deterg liq 500ml neutro ype, fica salientado o alcance da objetividade das regras de associação. Com participação de 0,19% no total das vendas, o produto **Massa C/ Ovos 500G Cas N.2 Dybom** aparece em Maio-Junho/09 com 47 transações. No período Maio-Junho/10 as operações com o mesmo item somaram 90 operações, o equivalente a 0,38% das vendas. Quando comparado o período Maio-Junho/09 e Maio-Junho/10 identifica-se aumento das vendas totais na ordem de 91%.

A tabela 28 indica uma variação positiva de 91% na venda do item massa c/ ovos 500g cas nº 2 dybom no período comparado. Evidenciando que alguns produtos podem não seguir os padrões de comercialização em um período parecido, chamando atenção para detalhes como a marca, o produto, local, e ou algum outro detalhe.

No período pesquisado, Maio-Junho/09 e Maio-Junho/10 o produto **Pêssego Calda 400G. Metades Cantu** obteve 2 e 110 transações respectivamente. No

primeiro período correspondeu a 0,01% do total de vendas e no segundo período respondeu por 0,46% das vendas totais, sugerindo elevação de 5.400% em termos de comercialização.

Nota-se que o item pêssego calda 400g. metades cantu teve um incremento de 5.400% quando associado ao item bata monaliza kg, e que se o item pêssego calda 400g. metades cantu permanecesse sem incentivo, a venda poderia seguir os números iniciais, queimando oportunidades de elevar vendas e por consequência rentabilidade do supermercado.

Muitas ações de *marketing* podem ser tomadas após o estudo dessas transações, auxiliando na reorganização das lojas, na produção de catálogos customizados, na seleção de produtos para promoção, na determinação de itens para propaganda, na seleção de artigos concorrentes, no fomento de *cross-selling*, e na identificação de oportunidades de vendas de pacotes de produtos ou serviços.

5 CONSIDERAÇÕES FINAIS

Neste capítulo serão expostas as conclusões gerais obtidas com a realização desta pesquisa.

Primeiro, com relação ao **suporte** das transações estudadas, é interessante a forma como o setor hortifrutigranjeiro aparece associado praticamente com todas as operações e com valores consideráveis, principalmente em dois produtos: batata monaliza e banana catura. Juntos eles somaram 3.818 transações, (ver tabela 1), indicando a importância do setor dentro do contexto geral. É possível que os consumidores, buscando uma melhor qualidade no consumo de alimentos, priorizem compras dentro do setor hortifrutigranjeiro. O supermercado em estudo possui uma área grande de exposição do setor hortifrutigranjeiro (ver fotos anexo⁸), com grande quantidade de produtos, conseguindo assim, atrair mais consumidores para consumo deste setor e por consequência do todo.

Segundo, na avaliação da variável **confiança** na regra de associação de $A \Rightarrow B$ e $B \Rightarrow A$ apesar de possuir suporte igual, percebe-se que as probabilidades são diferentes. Isso ocorre principalmente pelo grupo de interesse ou seja produto analisado. Na tabela 23, quando é comparada a associação dos produtos limp multi uso 500ml fta citr Q-bom \Rightarrow deterg liq 500ml neutro ype, tem-se uma confiança de 69%, ou seja, a cada cem compras de limp multi uso 500ml fta citr q-bom, é possível que o consumidor adquira 69 vezes deterg liq 500ml neutro ype. Quando os profissionais responsáveis pelo setor fazem análise das vendas, sejam elas, diárias, semanais, mensais e ou outros, eles conseguem perceber, através dos relatórios, as quantidades comercializadas dos produtos, porém, não conseguem visualizar como o consumidor os adquiriu.

Assim, o *Data Mining* pode descobrir índices de suporte e confiança no banco de dados que podem facilitar decisões do gestor, podendo-se concluir que as regras de associação tendem a facilitar o processo de tomada de decisão, fornecendo dados que vão auxiliar na exposição do mix de produtos no ponto de venda, bem como na melhoria do *layout* de distribuição dos produtos na gôndola.

⁸ O supermercado em estudo possui 100m² apenas para o setor de hortifrutigranjeiros.

Terceiro, a análise das tabelas **bimestrais** demonstra que os resultados encontrados têm forte relação com a tabela de análise anual, mesmo com a sazonalidade de determinados produtos, datas comemorativas, ou mesmo variações climáticas que podem sugerir preços diferentes do esperado para aquele momento. Em todas as análises bimestrais, nota-se os principais produtos vendidos: batata monaliza (kg) e banana catura (kg).

Quarto, foi realizado o estudo por suporte de **grupos isolados**, sendo: carne, material de limpeza, refrigerante, açúcar, óleo, sal, farinha de trigo e tomate. Com a realização da análise da regra $A \Rightarrow B$ identificou-se alguns itens que vendem associados a outros simplesmente. Assim, praticamente em todos os grupos de vinte produtos, os três maiores suportes estão associados ao setor hortifrutigranjeiro. Ainda, dos vinte produtos associados, praticamente 63% dos produtos pertencem ao setor hortifrutigranjeiro. Salientando, a importância deste setor dentro do supermercado.

Quinto, **os combos** sugeridos e implantados no supermercado indicam que pode-se ter períodos melhores para determinadas ofertas, como, por exemplo, início do mês quando a população economicamente ativa recebe o salário. Igualmente, períodos festivos do ano e feriados. Nos combos em teste, foi sugerido que fossem formados de acordo com a vontade do gestor, com exceção dos produtos batata monaliza (kg) e banana catura (kg), que indicavam números relevantes nas informações garimpadas dentro da ferramenta *Data Mining*. Nota-se, que os itens associados aos produtos batata monaliza (kg) e banana catura (kg), pêssego metades em calda cantu lata 400gr e doce de leite cantu 400gr, tiveram respectivamente um acréscimo de 5400% e 42,86%. Produtos como o Detergente Lava Louças Ype Neutro 500ml e Q-bon Multiuso Frutas Cítricas 500ml atingiram respectivamente percentuais de 121,52% e 8.400%, demonstrando ainda, que é possível associar produtos de diferentes setores e obter resultados significativos.

Sexto, foi possível suportar a teoria da aplicação de data mining em marketing, confirmando o impacto na estratégia de vendas da empresa varejista, de acordo com outros trabalhos, tais como Gonçalves (1999), Lamboia e Pereira (2005), e Baraniuk (2009). Os **resultados práticos** encontrados nesta pesquisa estão em consonância com os objetivos propostos naquelas investigações no que tange identificar as principais associações de compras.

Sétimo, no período anual percebe-se as **três principais associações**, sendo: batata monaliza kg \Rightarrow banana catura kg (3.818), tomate l.vida kg \Rightarrow batata monaliza kg (2.921), tomate l.vida kg \Rightarrow banana catura kg (2.866). Nota-se a importância da associação dos três itens garimpados, destacando o setor hortifrutigranjeiro em termos associativos.

Oitavo, através da variável suporte, no período bimestral, nota-se que a principal associação descoberta foi nos períodos: Jan até Fev, Mar até Abr, Mai até Jun, Jul até Ago, Set até Out e Nov até Dez/09. Dentro desses períodos a principal associação foi entre batata monaliza kg \Rightarrow banana catura, com suporte e confiança diversos.

Destarte, identifica-se que o comportamento das associações no períodos bimestrais tem alta correlação com os período anual, sugerindo que o comportamento das compras associadas segue um determinado padrão **temporal** no supermercado facilitando uma interpretação mais adequada do comportamento do consumidor quanto ao consumo de produtos associados e incentivando diversas possibilidades de combinações associativas.

Nas tabelas de grupo isolados, período pesquisado entre Jan até Dez/09, percebe-se através da variável suporte para seguintes associações, pernil suíno \Rightarrow banana catura kg (770), omo multi ação 1 \Rightarrow batata monaliza kg (1.312), coca pet 2lt \Rightarrow banana catura kg (1.704), açúcar cristal 5kg \Rightarrow batata monaliza kg (1.908), óleo cocamar 900ml \Rightarrow batata monaliza kg (1.606), sal moc 1 kg \Rightarrow batata monaliza kg (2.362), trigo koene 5kg \Rightarrow batata monaliza kg (1.109) e tomate l. vida kg \Rightarrow batata monaliza kg (2.921). As associações descobertas fortalecem os itens do setor hortifrutigranjeiro, principalmente batata monaliza kg e banana catura kg, ratificando o achado nas tabelas anuais e bimestrais.

Nono, os quatro combos sugeridos e implantados alusivos aos períodos comparados Maio/09 com Maio/10 demonstraram as seguintes variações,

- banana catura kg (2,28%)
- batata monaliza kg (9,37%)
- bisc salt 360g. tradicional águia 9,90%
- deterg liq 500ml neutro ype 72,73%
- doce 400g. leite cantu 65,93%
- limp multi uso 500ml fta citr q boa 628,57%

- massa c/ ovos 500g. cãs nº2 dybom (20%) e
- pêssogo calda 400g. metades cantu 770,59%

No período comparado Maio-Junho/09 e Maio-Junho/10, os seguintes produtos apresentaram as subseqüentes variações: banana catura kg (2,31%), batata monaliza kg 6,29% bisc salt 360g. tradicional água (22,73%), deterg liq 500ml neutro ype 121,52%, doce 400g. leite cantu 42,86%, limp multi uso 500ml fta citr q boa 8400%, massa c/ ovos 500g. cãs nº2 dybom 91,49% e pêssogo calda 400g. metades cantu 5.400%.

Com índices consideráveis nas variáveis suporte e confiança percebe-se que o consumidor adquire com frequência produtos em termos de estratégia de combos. Cabe aos gestores a responsabilidade de utilizar técnicas e ou métodos com origem na tecnologia da informação para transformar os dados armazenados em informações que possibilitem a interpretação de consumo. Além do mais, os combos permitiram confirmar que associando itens é possível elevar o nível de vendas.

5.1 Sugestões para Pesquisas Futuras

Dentro da pesquisa realizada, ficaram várias questões que podem vir a ser abordadas em pesquisas futuras. Destarte, algumas delas são apresentadas :

- Aplicar estudo semelhante em outros segmentos como farmácias, confecções, postos de combustíveis, tendo como premissa melhorar o conhecimento da ferramenta *Data Mining* e criar a possibilidade de comparar resultados alcançados;
- A partir da ferramenta *Data Mining*, aplicar diferentes técnicas de associação, como regras de associação multidimensionais, regra de associação híbrida, regras de associação multinível, regras de associação negativas, na busca de uma melhor interpretação do comportamento do consumidor;
- Pesquisar outras variáveis quanto ao impacto na associação como: os dias da semana e o turno (noturno/vespertino/matutino), a o interesse de vendas da organização de determinados produtos, ou mesmo testando a venda de itens combinados, aproximando produtos e necessidades, objetivando a facilidade de compra do consumidor;

- Descobrir perfil do consumidor que compra no supermercado (faixa etária, estado civil, etc.), classe social, forma de pagamento, direcionando ofertas do tipo combo por perfil sócio demográfico;
- Criar e mensurar combos de vendas com quatro ou mais itens com objetivo de alavancar estratégias de vendas;
- Pesquisar como as marcas podem influenciar e facilitar as associações de compras (Johnson e Johnson e Procter & Gamble).
- Análise de data mining por produtos próximos ao ponto-de-venda (nas proximidades do caixa), realizando um mapeamento por regiões do mercado.

5.2 Sugestões Práticas para o Supermercado

Através da pesquisa realizada é possível relacionar algumas sugestões almejando melhorias que podem ser úteis ao dia-a-dia do supermercado, sendo elas:

- Reavaliar forma como os produtos estão sendo codificados no sistema. As vezes o código principal é o código sugerido pelo sistema e, por vezes, aparece outro no código de barras, dificultando análises com ferramentas como *Data Mining*;
- Procurar não sobrepor itens na base de dados, quando do ato de realizar cadastros de novos produtos e ou fornecedores. Por exemplo, quando for cadastrar um item, não sobrepor outro item (código de produto) que saiu de linha;
- Quando for destacar uma oferta associada com dois ou mais itens, por exemplo, sinalizar a economia que o consumidor esta realizando com aquela aquisição; (adquirir item A por R\$ 1,00 + item B por R\$ 1,00, juntos apenas R\$ 1,80 – economize R\$ 0,20).
- Implantar cartão fidelidade aos clientes, premiando-os de acordo com as pontuações, mapeando suas compras;
- Reposicionar o setor hortifrutigranjeiro dentro do layout do supermercado, melhorando exposição/exibição dos produtos. Os resultados da pesquisa indicam que este setor é de extrema relevância dentro do mix ofertado no supermercado.

- Inverter o local do setor de bebidas com o setor de farinhas e grãos, criando a necessidade de maior circulação dentro do supermercado. Da forma que se encontra o layout atual, o consumidor que for adquirir farinhas e ou grãos não tem necessidade de se locomover dentro do supermercado, perdendo a oportunidade de venda por impulso por exemplo.
- Comercializar produtos eletrônicos como celulares, informática e outros.

5.3 Limitações do Trabalho

Este trabalho teve como principal obstáculo o desenvolvimento de uma ferramenta com capacidade de extrair as informações do banco de dados do supermercado de forma a não gerar prejuízos a leitura dos dados. O principal motivo encontrado foi que os cadastros dos produtos não seguiram um padrão, ora tendo como código a sequência do *software* e ora o código de barras, gerando confusão na extração dos dados. Assim, os dados brutos, uma vez que estavam cadastrados errados na base do supermercado, podem gerar algum limite do resultado da pesquisa.

A sobreposição no cadastro de produtos descontinuados também indicou preocupação ao pesquisador para que não realizasse a pesquisa de um item com um determinado código. Isto ocorreu pelo fato de que duplicações nos códigos dos produtos dificultassem a elaboração de uma base de dados mais fidedigna. Esta é a segunda limitação do trabalho.

Terceiro, os combos sugeridos e implantados estavam associados em dois itens. A ferramenta *data mining* devolveu a informação demonstrando que é possível trabalhar com uma quantidade maior de itens, por exemplo 3 ou 4, inclusive impulsionando as variáveis suporte e confiança. Assim, as sugestões de 2 produtos é limitada e pode ser mais explorada em pesquisas futuras, introduzindo 3 itens, como é o caso do McDonalds (lanche+refrigerante+batata).

Quarto, o tempo que o combo ficou exposto é um limitante. Embora o acompanhamento foi feito em apenas 1 mês, torna-se necessário uma exposição maior do combo para o cliente final.

Quinto, a base de dados é limitada a apenas 1 supermercado da região sudoeste. Utilizar grandes bases de dados para checar preferências e associações de compra podem provocar resultados distintos e interessantes para a estratégia de marketing.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. **Fast Algorithms for Mining Association Rules**. In: 20th INTERNACIONAL CONFERENCE ON VERY LARGE DATABASE. Santiago del Chile: Morgan Kaufmann, 1994, p. 487-499.

AGRAWAL R, Imielinski T, and Swami. **A Mining association rules between sets of items in large databases**. In Proceedings... The ACM SIGMOD International Conference on Management of Data, 1993, p. 207–216.

ALMEIDA, F. C. **Desvendando o uso de redes neurais em problemas de administração de empresas**. São Paulo: Revista de Administração de Empresas - RAE. v. 35, n.1, Jan./Fev. 1995. p.46-55.

ANDRADE, Maria Margarida de. **Como preparar trabalhos para cursos de pós-graduação: noções práticas**. 5. Ed. São Paulo: Atlas, 2002.

ANGELONI, M. T. **Elementos intervenientes na tomada de decisão**. Brasília: Ciência da Informação. vol. 32 n.1, Jan./Apr. 2003, p.18.

ARAUJO, O. **O composto mercadológico do varejo**. 2010. Disponível em: <<http://www.dearaujo.ecn.br/cgiin/asp/compostoMercadologicoVarejo.asp>>. Acesso em: 17 abr. 2010. p.1.

ASSUNÇÃO, L. **Data Warehousing**. 2001. Disponível em: <http://utilsoft.com.br/help/index.html?data_warehousing.htm>. Acesso em: 23 jan. 2010. p. 1.

ATIKNSON, R.; DILL, S. L. **Como organizar um banco de dados moderno e eficiente**. In: Celepar. 2009. Disponível em: <<http://www.batebyte.pr.gov.br/modules/conteudo/conteudo.php?conteudo=909>> Acesso em: 17 abr. 2010.

BARANIUK, J. A. **Modelo de Gerenciamento de Serviços, Utilizando o Valor do Cliente no Tempo: Uso de Mineração de Dados em um Serviço de Telecomunicações**. Tese de Doutorado. Universidade Federal do Paraná. 2009. p. 1 a 260.

BARBOSA, A. L. S. **Como a Tecnologia da Informação, através da Inteligência Empresarial – “Business Intelligence” (BI), tem provocado impacto na gestão**

das empresas? 2007. Disponível em:
[http://www.artigos.com/artigos/sociais/administracao/como-a-tecnologia-da-informacao,-atraves-da-inteligencia-empresarial-%96-%93business-intelligence%94-\(-bi-\),--tem-provocado--impacto-na-gestao-das-empresas?-2107/artigo/](http://www.artigos.com/artigos/sociais/administracao/como-a-tecnologia-da-informacao,-atraves-da-inteligencia-empresarial-%96-%93business-intelligence%94-(-bi-),--tem-provocado--impacto-na-gestao-das-empresas?-2107/artigo/)>. Acesso em: 23 jan. 2010. p.1.

BERRY, L.L., ZEITHAML, V. and PARASURAMAN, A. **Quality counts in service, too, Business Horizons**, May-June, 1985, p. 45–6.

BERRY, Michael J. A.; LINOFF, Gordon. **Data Mining Techniques: For Marketing, Sales, and Customer Support**. New York: Wiley Computer Publishing, 1997.

BONOMO, P. **Construção de Data Warehouse (DW) e Data Mart (DM)**. Imasters. 2009. Disponível em:
http://imasters.uol.com.br/artigo/11178/gerencia/construcao_de_data_warehouse_dw_e_data_mart_dm/>. Acesso em: 17 abr. 2010. p. 1 a 3.

CABRINO, T. **Marketing de Relacionamento**. Portal do Marketing. 2002. Disponível em:
<http://www.portaldomarketing.com.br/Artigos/Marketing%20de%20Relacionamento.htm>>. Acesso em: 23 jan. 2010.

CAMPOS, O. B. **Data Mining**: overview. Disponível em:
<http://www.de9.ime.eb.br/~intec/Data%20Mining/Artigos%20de%20Suporte/Overview%20Data%20Mining.pdf>>. Acesso em: 24 Abr. 2010.

CLARO, D. P. **Marketing de relacionamento**: conceitos e desafios para o sucesso do negócio. São Paulo: lbmec, 2006.

CÔRTEZ, P. L. **Gestão da Tecnologia da Informação**. Catho e-learning. 2004, p. 122 e 137.

DIAS, S. **Banco de dados**. 2008. Disponível em:
<http://www.administradores.com.br/informe-se/artigos/banco-de-dados/39764/>>. Acesso em: 25 abr. 2010.

ESTENDER, A. C. **O novo varejo**. Instituto Siegen. Disponível em:
http://www.institutosiegen.com.br/artigos/ARTIGO_ESTENDER01.pdf>. Acesso em: 17 abr. 2010.

EXAME. **As 500 maiores e melhores empresas do Brasil**. São Paulo: Abril, jun. 2000.

FINGERL, E. R. **Comércio Varejista: Supermercados**. Rio de Janeiro. 1996. Disponível em: http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/relato/supmerca.pdf. Acesso em: 17 abr. 2010.

GIANNOTTI, F.; MANCO, G.; NANNI, M.; PEDRESCHI, D.; TURINI, F. **Integration of Deduction and Induction for Mining Supermarket Sales Data**. In: Proceeding of the Third International Conference on Practical Applications of Knowledge Discovery and *Data Mining* (PADD'99), sep.1999, p.1.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas 2002.

GIL, A.C. **Métodos e técnica de pesquisa social**. 5. Ed. São Paulo Atlas, 1999.

GOMES, Isabela M. **Manual Como Elaborar uma Pesquisa de Mercado**. Belo Horizonte: SEBRAE/MG, 2005. p.20.

GONÇALVES, L. P. F. **Mineração de Dados em Supermercados: O caso do supermercado "Tal"**. Porto Alegre: UFRGS-URCAMP. 1999, p.3 a 26.

GONÇALVES, E. C. **Data Mining de Regras de Associação – Parte 1**. Disponível em: <http://www.devmedia.com.br/articles/post-6533-Data-Mining-de-Regras-de-Associacao-Parte-1.html>. Acesso em: 27 Abr. 2010.

GUPTA, S.; LEHMANN, D. R. Customers as assets. **Journal of Interactive Marketing**, v. 17, n. 1, 2003, p. 10.

HAN, J.; Micheline K. **Data Mining: Concepts and Techniques**. Academic Press, Morgan Kaufmarm Publishers, 2001.

HARRISON, T. H. **Intranet Data Warehouse**. São Paulo: Bekerley Brasil, 1998.

HIDBER, C. **Online association rule mining**. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA. Philadelphia: SIGMOD, 1999, p. 145-156.

LAMBOIA, F.; PEREIRA, L. M. **Data Mining**. Cascavel: Monografia Unioeste. 2005, p. 5 a 20.

LAS CASAS, A. L.; BARBOZA, V. A. **Marketing no Varejo**. In: ESTRATÉGIAS DO MARKETING PARA VAREJO: Inovações e diferenciações estratégicas que fazem a diferença no marketing de varejo. Novatec. 2007. p. 5 a 23.

LESCA, H.; ALMEIDA, F.C.; **Administração Estratégica da Informação**. Revista de Administração. São Paulo. v.29. n.3, p.66, jul/set. 1994.

LEXTEC. **Supermercado**. Disponível em: <http://www.instituto-camoes.pt/lextec/por/domain_3/definition/4946.html>. Acesso em: 17 abr. 2010. p.1.

LOVIN, Radu <http://www.dataminingarticles.com/association-rules.html>. **Easy Data Mining - Part 3**. Acesso em 03 de Maio de 2010. p.1

MENON, S.; Sharda, R. **Data Mining update: new models to pursue old objectives**, OR/MS Today, 26(3), 1999, p. 26–9.

Min H. 2006. **Developing the profiles of supermarket customers through data mining**. *Service Industries Journal* 26(7), pp. 747-763.

MINAYO, M. C. S. (Org.). **Pesquisa Social** – teoria, método e criatividade. Rio de Janeiro: Vozes, 2001.

NAVEGA, S. C. **Princípios Essenciais do Data Mining**. Intellwise Research and Training. Agosto, 2002, p.2.

PARENTE, J. **Varejo no Brasil**. São Paulo: Atlas, 2000. p. 22.

PASSARI, A. F. L. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais**. São Paulo: USP. 2003. p. 7 a 46.

PASSINI, S. R. R.; TOLEDO, C. M. T.; **Mineração de Dados para detecção de fraudes em ligações de água**. XI Seminco: Seminário de Computação. 2002.

PERINI, L. C.; RAIMUNDO, C. M. **Um estudo exploratório sobre o papel da Informação na tomada de decisão para o administrador**. 2009. Disponível em: <<http://www.artigonal.com/administracao-artigos/um-estudo-exploratorio-sobre-o-papel-da-informacao-na-tomada-de-decisao-para-o-administrador-852812.html>>. Acesso em: 23 jan. 2010.

PIATETSKY-SHAPIO, G. . **Discovery, analysis and presentation of strong rules**. In Knowledge Discovery in Databases. AAI/MIT-Press, 1991.

PINHO, A. G. **Mineração de Dados com Mapas de Kohonen**: Uma abordagem no Setor Financeiro. XXXIII Encontro ANPAD. Set. 2009.

RAUPP, Fabiano Maury; BEUREN, Ilse Maria. Metodologia da pesquisa aplicável nas Ciências Sociais. In: BEUREN, Ilse Maria (Org.). **Como elaborar trabalhos monográficos em contabilidade**: teoria e prática. São Paulo: Atlas, 2003. p.81.

REZENDE, S. O., OLIVEIRA, R. B. T.; FELIX, L. C. M. & ROCHA, C. A. J. **Visualization for knowledge discovery in database**. In *Ebecken, N.F.F. (ed.) Data Mining*. WIT Press, England, 1988. pp. 81–95.

Rezende, S. O., Pugliesi, J. B., Melanda, E. A., and de Paula, M. F. (2003). **Mineração de dados**. In Rezende, S. O., editor, *Sistemas Inteligentes: Fundamentos e Aplicações*, pages 307–335, Barueri, SP. Manole.

RICHTER, H. G. **Retailing**: principles and practices. New York: Mc Graw Hill, 1954. p. 53.

SACCOL, A. Z., PEDRON, C. D., NETO, G. L., MACADAR, M. A., CAZELLA, S. C., **Avaliação do Impacto dos Sistemas ERP sobre Variáveis Estratégicas de Grandes Empresas no Brasil**. Revista de Administração Contemporânea, v. 8, n. 1, Jan./Mar. 2004. p. 13.

SANTANA, A. G.; FREITAS, G. A. **Gerenciamento de TI**. Linhares: UNILINHARES. 2006. p. 7 a 31.

SANTOS, C. B.; ZANDER, V. T.; SCANDELARI, L.; ROMANO, C. A.; CARVALHO, D. R. **Data Mining aplicado na base de dados das categorias mais vendidas de um supermercado**. XXVI ENEGEP – Fortaleza, CE, Brasil, 9 a 11 de Outubro de 2006.

SEBRAE-ES. **Análise de negócios- Supermercados e mercearias**. Disponível em: <
[http://201.2.114.147/bds/BDS.nsf/485CCBAE36839CEC832573BE00520FD8/\\$File/Supermercados_e_Mercearias.pdf](http://201.2.114.147/bds/BDS.nsf/485CCBAE36839CEC832573BE00520FD8/$File/Supermercados_e_Mercearias.pdf)>. Acesso em: 17 abr. 2010. p. 1.

SEIXAS, T. S. **Análise CLV**. Cadeira e-marketing. Universidade da Madeira. 2007. p. 4.

SERAIN, J. S. **O BI e a análise do ambiente**. Disponível em: <
http://imasters.uol.com.br/artigo/6988/bi/o_bi_e_a_analise_do_ambiente/imprimir/>. Acesso em: 17 abr.2010.

SILVEIRO NETTO, C. B. **O que é o ERP?** CBS Consultoria. Disponível em: <
<http://www.cbsconsulting.com.br/erp.htm>>. Acesso em: 17 abr. 2010.

SOUZA, C. A.; ZWICKER, R. **Gestão de Sistemas ERP**: Uma Análise das capacidades e atores envolvidos. 30º. Encontro ANPAD. Salvador. Set/2006. p. 3.

STUMPF, I. R. C. Pesquisa bibliográfica. In: **Métodos e técnicas de pesquisa em Comunicação**. São Paulo: Atlas, 2005. p. 51 a 61.

SUMATH, S.; SIVANANDAM, S. N. **Introduction to Data Mining and its applications**. Berlin: Springer-Verlag, 2006.

VIEIRA, V. A.. **As tipologias, variações e características da pesquisa de marketing**. Revista da FAE. Curitiba. v.5, n.1, p.61 – 70, jan./abr. 2002.

WILSON, J. **Sistema de gerenciamento de banco de dados**: Introdução a SGDB. 2010. Disponível em: <
www.cepceilandia.df.gov.br/sites/200/296/00000485.ppt>. Acesso em: 17 abr. 2010. p.5.